AFRL-RH-WP-TR-2007-0123

Grounding Symbolic Operations in Modality-Specific Processing

Lawrence W. Barsalou
Aron Barbey
Joy Lynn Brasfield
Shlomit Finkelstein
Andrew James
Xiaoping Hu
Zhaohui Liu
Ava Santos
Christine Wilson

Department of Psychology
Emory University
Atlanta GA 30329

Stefan Gries

Department of Psychology
University of California
Santa Barbara CA 93106

Alex Martin
Kyle Simmons

National Institute of Mental Health
6001 Executive Boulevard
Rockville MD 20852

July 2007

Final Report for July 2005 to July 2007

Approved for public release; distribution is unlimited.

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
Cognitive Systems Branch
Wright-Patterson AFB OH 45433-7022

NOTICE

AFRL-RH-WP-TR-2007-0123

//SIGNED//
JOHN L. CAMP
Work Unit Manager
Cognitive Systems Branch

//SIGNED//
DANIEL G. GODDARD
Chief, Warfighter Interface Division
Human Effectiveness Directorate
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
July 2007	Final	July 2005 – July 2007

4. TITLE AND SUBTITLE
Grounding Symbolic Operations in Modality-Specific Processing

5a. CONTRACT NUMBER FA8650-05-C-7256

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

[1]Lawrence W. Barsalou, [1]Aron Barbey, [1]Joy Lynn Brasfield, [1]Shlomit Finkelstein [1]Andrew James, [1]Xiaoping Hu, [1]Zhaohui Liu, [1]Ava Santos, [1]Christine Wilson, [2]Stefan Gries, [3]Alex Martin, [3]Kyle Simmons

5d. PROJECT NUMBER 63123F

5e. TASK NUMBER

5f. WORK UNIT NUMBER DRPA0508

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) ND ADDRESS(ES)

[1]Department of Psychology
Emory University
Atlanta GA 30329

[2]Department of Psychology
University of California
Santa Barbara CA 93106

[3]National Institute of Mental Health
6001 Executive Boulevard
Rockville MD 20852

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
Cognitive Systems Branch
Wright-Patterson AFB OH 45433-7022

10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RHCS

11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-RH-WP-TR-2007-0123

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

88TH ABW/PA cleared on 12 December 2007, WPAFB-07-0677.

14. ABSTRACT

The research performed under this contract assessed whether fundamental symbolic operations—predication, conceptual combination, and the representation of abstract concepts—arise from the simulation of modality-specific states in the brain. Traditionally, symbolic operations have been widely assumed to arise from the manipulation of amodal symbols. Indeed, researchers often assume that symbolic operations could *only* result in this latter way. Recent research on grounded cognition, however, has proposed that symbolic operations, in principle, could arise from modality-specific simulation. The experiments performed here offer preliminary evidence that they do.

15. SUBJECT TERMS Cognitive Architecture, Collaboration, Communication, Decision Making, Modeling & Simulation, Human Performance, Human-Computer Interface

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					John L. Camp
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR	66	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

THIS PAGE LEFT INTENTIONALLY BLANK

TABLE OF CONTENTS

List of Figures

List of Tables

I. Background and Goals

The research performed under this contract assessed whether fundamental symbolic operations—predication, conceptual combination, and the representation of abstract concepts—arise from the simulation of modality-specific states in the brain. Traditionally, symbolic operations have been widely assumed to arise from the manipulation of amodal symbols. Indeed, researchers often assume that symbolic operations could *only* result in this latter way. Recent research on grounded cognition, however, has proposed that symbolic operations, in principle, could arise from modality-specific simulation. The experiments performed here offer preliminary evidence that they do.

These findings inform attempts to build computational agents that perform end-to-end processing during situated action in an environment. To function effectively, such agents must acquire categorical knowledge of objects, events, mental states, etc., and they must manipulate this knowledge symbolically, using fundamental cognitive operations such as predication and conceptual combination. Furthermore, to understand their own mental states and how they relate to events in the world, a computational agent must be able to represent abstract concepts. The experiments performed here explore simple paradigms like those that face computational agents in their simple environments, and offer guidance in designing their computational architectures.

None of the six projects performed here used a previously established paradigm. Instead, each project developed a new paradigm that either addressed new issues or that addressed an established issue in a new way, often with the aim of assessing modality-specific simulation. All these new paradigms offer new tools for exploring the roles of modality-specific simulation in cognition. Two projects also developed technical procedures not previously used before.

Before presenting the results of our research, we provide further background on cognitive architecture and symbolic operations. We then present each project, first describing its methods and the innovations they offer. We then present results from the project and their implications.

A. Cognitive Architectures

Figure 1 illustrates the standard cognitive architecture that underlies widespread thinking about the representation of knowledge. Figure 2 illustrates an alternative architecture that underlies recent embodied views. Depending on the architecture that a researcher adopts, different ways of thinking about symbolic operations follow. Each architecture is addressed in turn.

1. The transduction of amodal symbols in standard cognitive architectures. Standard architectures assume that amodal symbols are transduced from experience to represent knowledge. Figure 1 illustrates this general approach. On experiencing a member of a category (e.g., *dogs*), modality-specific states arise in the visual system (the black arrows in Panel A), auditory system (orange arrows), motor system (blue arrows), somatosensory system (purple arrows), etc. These states represent sensory-motor information about the perceived category member, with some (but not all) of this information producing conscious experience. Although modality-specific states are shown only for sensory-motor systems, we assume that modality-specific states also arise in motivational systems, affective systems, and cognitive systems. We will refer to the perception of these internal systems as *interoception* from here on. Once modality-specific states arise in all relevant modality-specific systems for a category, amodal symbols that stand for conceptual content in these states are then transduced elsewhere in the brain to represent knowledge about the category (e.g., *legs, tail, barks, pat, soft* in Panel B for the experience of a dog). Although words often stand for transduced amodal symbols (e.g., *leg*), most theories assume that sub-linguistic symbols, often corresponding closely to words, are actually the symbols transduced (e.g., § in Panel B).

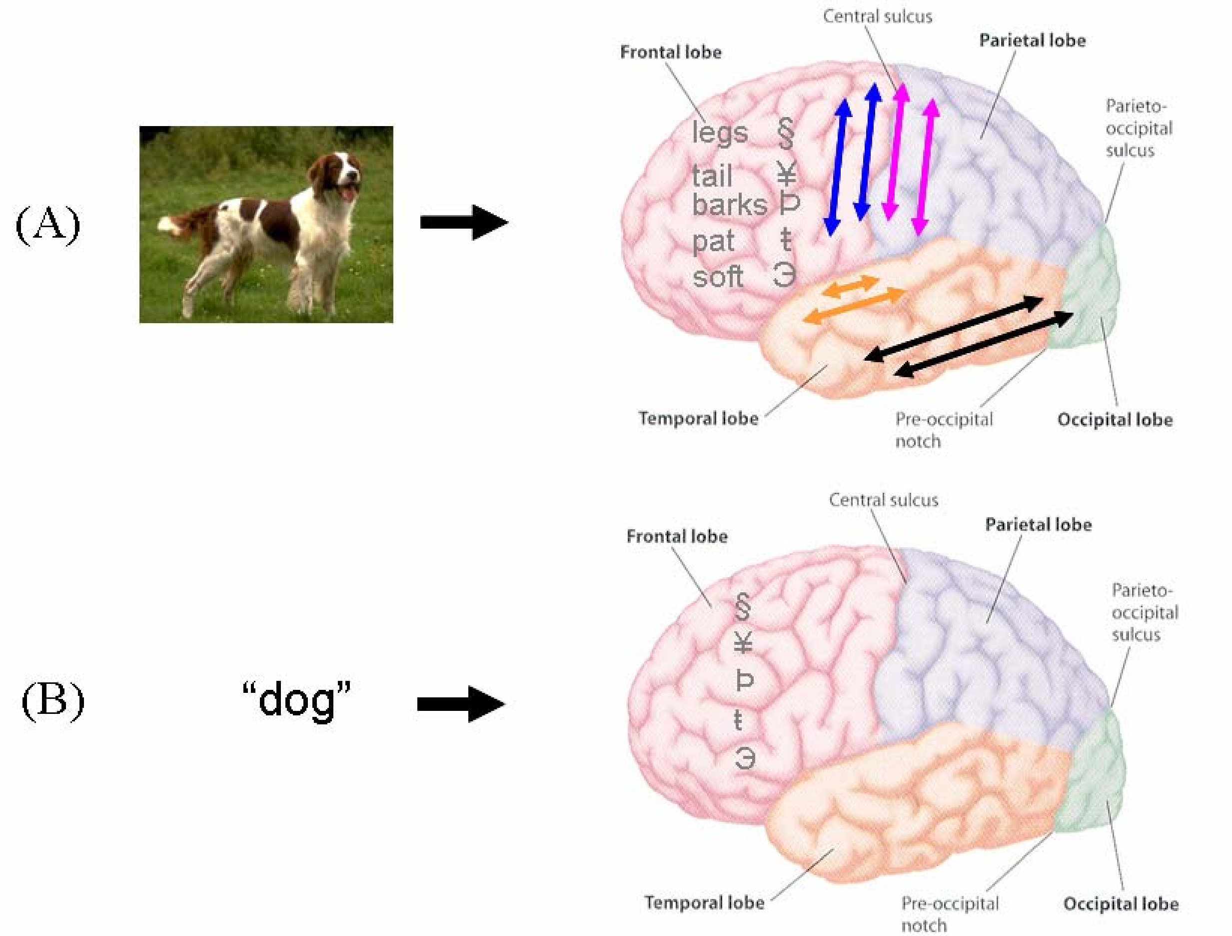


Figure 1. The transduction of amodal symbols from modality-specific states in standard cognitive architectures (Panel A). Use of transduced symbols to represent the meaning of a word (Panel B). See the text for further description.

Once established in the brain, amodal symbols later represent knowledge about the category across a wide range of cognitive tasks (Figure 1, Panel B). During language comprehension, for example, hearing the word for a category (e.g., “dog”) activates amodal symbols transduced from modality-specific states on previous occasions. Subsequent cognitive operations on category knowledge, such as inference, are assumed to operate on these symbols. Note that none of the modality-specific states originally active when amodal symbols were transduced are active during knowledge representation. Instead, amodal symbols are assumed to be sufficient and modality-specific states irrelevant.

The architecture in Figure 1 underlies a wide variety of standard approaches to representing knowledge, such as feature lists, semantic networks, and frames. This architecture also underlies those neural net architectures where the hidden layers that represent knowledge are related arbitrarily to perceptual input layers. This architecture does not underlie neural net architectures where input units play roles in knowledge representation.

2. The capture and simulation of modality-specific states in grounded cognitive architectures. A very different approach to representing knowledge has arisen recently in grounded theories of cognition (Barsalou, 2008). Actually, this approach has deep roots in philosophical treatments of knowledge going back over 2000 years (e.g., Barsalou, 1999; Prinz, 2002). Modern theories can be viewed as reinventions of these older theories in the contexts of psychology, cognitive science, and cognitive neuroscience. Interestingly, the amodal architectures that currently dominate the field constitute a relatively recent and short presence in a historical context where theories grounded in the modalities have dominated.

Figure 2 illustrates the grounded approach to representing knowledge. On experiencing a member of a category (e.g., *dogs*), modality-specific states are represented as activations in the

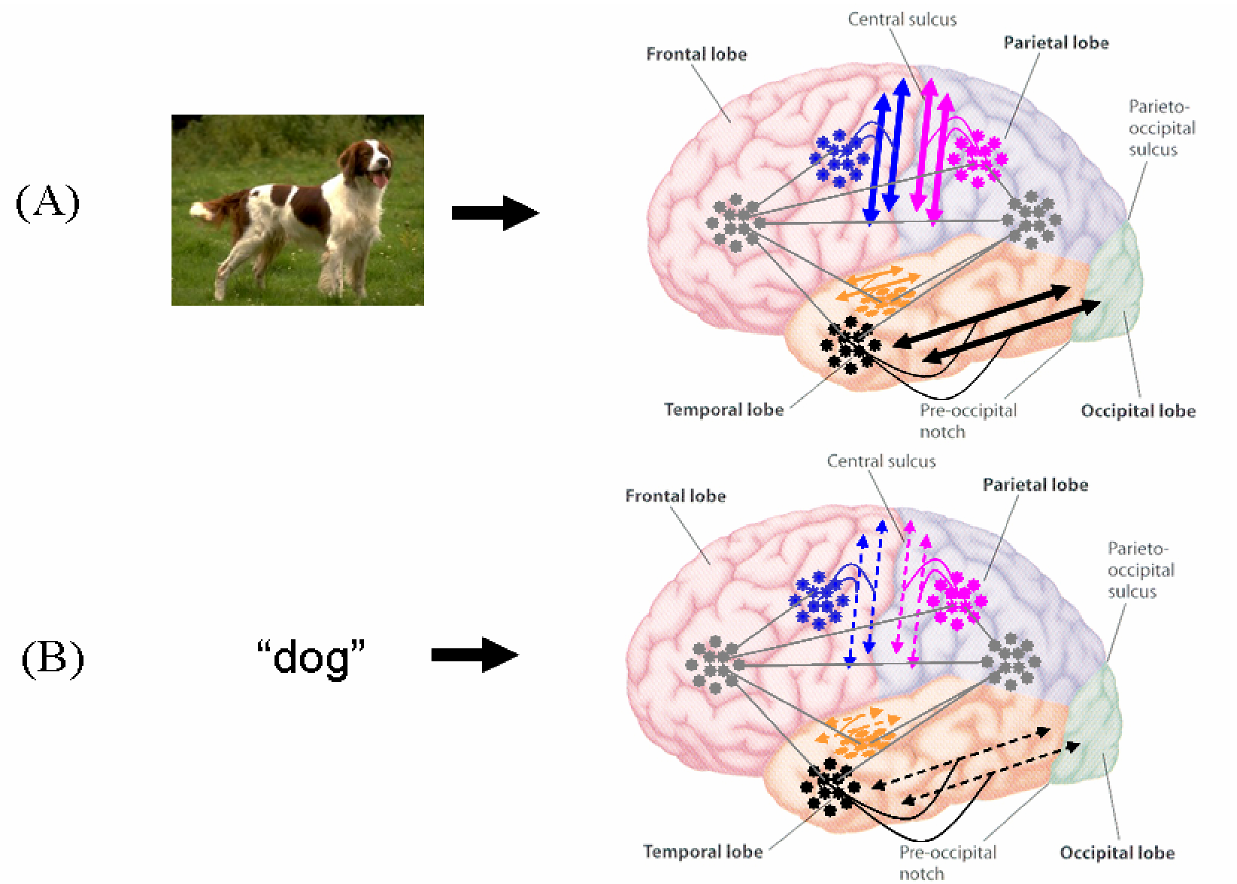


Figure 2. Conjunctive units in hierarchically-organized association areas capture modality-specific states across modalities in grounded theories of knowledge (Panel A). Captured multi-modal states are simulated to represent the meaning of a word (Panel B). See the text for further description.

visual system (the black arrows in Panel A), auditory system (orange arrows), motor system (blue arrows), somatosensory system (purple arrows), etc. As for Figure 1, modality-specific states are only shown for sensory-motor systems, but we assume that such states are also captured during the interoception of motivational systems, affective systems, and cognitive systems. Local association areas then partially capture these modality specific states (shown in Panel A as stars in the same color as the captured states). Higher-order cross-modal associations (gray stars) then integrate conjunctive neurons in lower-order association areas to establish a multi-modal representation of the experience.

Once established in the brain, these multi-modal associative structures represent knowledge about the category across a wide range of cognitive tasks (Figure 2, Panel B). During language comprehension, for example, hearing the word for a category (e.g., "dog") activates conjunctive neurons in higher-order cross-model association areas that have previously encoded experiences of the respective category. In turn, these conjunctive neurons activate lower-order conjunctive neurons that partially reactivate modality-specific states experienced previously for the category. These neural reenactments attempt to simulate the modality-specific states likely to occur when encountering category members. See Damasio (1989) and Simmons and Barsalou (2003) for further detail.

The architecture in Figure 2 underlies a wide variety of traditional and modern approaches to representing knowledge. Whereas some of these approaches focus on mental images, others focus on neural reenactments of modality-specific states in the brain. All share the common assumption that the representation of knowledge is grounded in modality-specific states, rather than in amodal symbols transduced from them.

B. The Status of Empirical Evidence for Grounded Knowledge

Accumulating empirical evidence supports the simulation architecture in Figure 2. Many findings indicate that the brain's modality-specific systems for perception, action, and interoception are active during the higher cognitive activities of memory, knowledge, language, and thought. For reviews of evidence from cognitive psychology, see Barsalou (2003b) and Barsalou, Simmons, Barbey, and Wilson (2003). For reviews of evidence from social psychology, see Barsalou, Niedenthal, Barbey, and Rupport (2003) and Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric (2005). For reviews of evidence from cognitive neuroscience, see Martin (2001, 2007), Pulvermüller (1999), and Thompson-Schill (2003). For reviews of developmental evidence, see Smith and Gasser (2005) and Thelen (2000). For a general review across areas, see Barsalou (2008). The rapidly accumulating findings across these diverse literatures indicate that the higher cognitive processes engage modality-specific systems frequently and robustly.

Problematically, however, these findings do not indicate what roles the modalities play. When these findings were acquired, it was not accepted widely that modality-specific systems participated in higher cognition at all. Researchers holding this hypothesis therefore attempted to evaluate it primarily using demonstration experiments. These researchers did not attempt to establish the roles that modality-specific processing played in the experimental phenomena studied. Now that the presence of modality-specific processing is becoming well established, however, demonstration experiments are likely to have diminishing returns. Instead, it is becoming increasingly important to establish the specific roles that the modalities play.

One possibility is that the brain's modality-specific systems play relatively peripheral, or even epiphenomenal, roles in higher cognition. Although these systems become active, other systems that operate on amodal symbols implement basic cognitive operations.

Alternatively, the theory of Perceptual Symbol Systems (PSS) proposes that the brain's modality-specific systems provide the core computational engine in higher cognition (Barsalou, 1999; 2003a, 2005). In particular, PSS proposes that simulators and simulations grounded in modality-specific systems implement fundamental symbolic operations, such as binding types to tokens, binding arguments to values, drawing inductive inferences from category knowledge, predicating properties and relations of individuals, combining symbols to form complex symbolic expressions, and representing abstract concepts that interpret meta-cognitive states. The research performed in this project evaluated whether symbolic operations like these are grounded in the brain's modality-specific systems. For a review of current evidence showing that symbolic operations are grounded in the modalities, see Barsalou (in press).

C. Symbolic Operations

A central theme of modern cognitive science is that symbolic interpretation underlies human intelligence. The human brain does not simply register images, as do cameras and other recording devices. A collection of images or recordings does not make a system intelligent. Instead symbolic interpretation of image content is essential for capturing the intelligent activity of biological agents and for implementing it in intelligent ones.

What cognitive operations underlie symbolic interpretation? Across decades of analysis, a consistent set of symbolic operations has arisen repeatedly in logic and knowledge engineering: binding types to tokens; binding arguments to values; drawing inductive inferences from category knowledge; predicating properties and relations of individuals; combining symbols to form complex symbolic expressions; representing abstract concepts that interpret meta-cognitive states. It is difficult to imagine performing intelligent computation without these operations. For this reason, many theorists have argued that they are central not only to artificial intelligence, but to human intelligence (e.g., Fodor, 1975; Pylyshyn, 1973).

Symbolic operations provide an intelligent system with considerable power for interpreting its experience. Using type-token binding, an intelligent system can place individual components of an image into familiar categories (e.g., categorizing components of an image as people and cars). Rich inferential knowledge then results from retrieving information from these categories that allows the perceiver to predict how categorized individuals will behave, and to select effective actions that can be taken (e.g., a perceived person is likely to talk, cars can be driven). Symbolic knowledge further allows a perceiver to predicate properties about the individual that may be useful to pursuing relevant goals with it (e.g., predicating that an object is likely to explode). Such predications further support high-level cognitive operations, such as decision making (e.g., does a person have the properties of a terrorist), planning (e.g., if a car contains a bomb, what actions might prevent explosion), and problem solving (e.g., how can a stalled car be moved). Symbolic operations include a variety of operations for combining symbols, such that an intelligent system can construct complex symbolic expressions (e.g., combining word meanings during language comprehension). Finally, by establishing abstract concepts about mental states and operations, an intelligent system can categorize its mental events, and can reason about how to manipulate them (e.g., constructing and monitoring a plan for driving to a destination).

1. Possible accounts of symbolic operations. What mechanisms implement symbolic operations? Since the cognitive revolution, language-like symbols and operations have been widely assumed to make these operations possible. As a result, numerous theoretical approaches have been grounded in predicate calculus and propositional logic. Not only have these approaches been central in artificial intelligence (e.g., Charniak & McDermott, 1985), they have also been central throughout accounts of human cognition (e.g., Anderson, 1983; Newell, 1990; Pylyshyn, 1984).

Although classic symbolic approaches are still widely accepted as accounts of human intelligence, and also as the engine for artificial intelligence, they have come increasingly under attack for two reasons. First, classic symbolic approaches have been widely criticized for not being sufficiently statistical. As a result, neural net approaches have developed to remedy this deficiency (e.g., Rumelhart & McClelland, 1986; O'Reilly & Munakata, 2000). Second, classic symbolic approaches have been criticized for not being grounded in perception, action, and interoception. As a result, researchers have argued that higher-order cognition is grounded in the brain's modality-specific systems. Although few computational models for this latter approach exist yet, empirical support has become quite strong (e.g., Barsalou, 2003b, 2008; Barsalou, Simmons, et al., 2003; Barsalou, Niedenthal et al., 2003; Martin, 2001; Niedenthal et al., 2005; Smith, 2005; Thelen, 2000; Thompson-Schill, 2003).

As statistical and grounded approaches become increasingly embraced, the tendency to throw the baby out with the bath water has arisen. Some researchers have suggested that classic symbolic operations are irrelevant to higher cognition, especially researchers associated with neural nets and dynamical systems (e.g., van Gelder, 1990). Notably, however, some neural net researchers have realized that symbolic operations are essential for implementing higher cognitive phenomena in knowledge, language, and thought. The problem in classic theories is not their inclusion of symbolic operations, but *how* they implement them. For this reason, neural net researchers have developed neural net accounts of symbolic operations (e.g., Pollack, 1990; Smolensky, 1990). Interestingly, however, these approaches have not caught on widely, either with psychologists, or with knowledge engineers. For psychologists, neural net accounts of symbolic processes have little psychological plausibility; for knowledge engineers, they are difficult and inefficient to implement. As a result both groups continue to use classic theoretical frameworks when symbolic operations must be implemented.

An alternative account of symbolic operations has arisen in grounded theories (Barsalou, 1999, 2003a, 2005). Not only does this account have psychological and neural plausibility, it suggests a new approach to image analysis. Essentially, this approach develops symbols whose content is extracted from images. As a result, image-based symbols can be bound to the regions of other images, thereby establishing type-token mappings without using amodal symbols. Inferences drawn from category knowledge also take the form of images, such that they can be mapped to perception. Analysis of an individual in an image proceeds by processing its perceived regions and assessing whether perceptually grounded properties and relations can be predicated of them. Symbol combination involves the manipulation and integration of image components to construct structured images that, in effect, implement complex symbolic propositions. Abstract concepts result from interoception, namely, the process of perceiving meta-cognitive states and developing image-based representations of them for later use in reasoning. This approach offers an exciting new way of thinking about the symbolic operations that underlie intelligence. It also offers a powerful way of interfacing higher cognition with perception, action, and interoception. The following sub-sections present how PSS explains symbolic operation in further detail.

2. Simulators and simulations. To implement symbolic operations, it is essential for an intelligent system to have some means of learning and representing concepts. The lack of concepts is what prevents many recording devices, such as cameras and video recorders, from implementing the symbolic operations that would allow them to interpret the images they capture. The central innovation of PSS (Perceptual Symbol Systems) is its ability to implement concepts using image content as basic building blocks.

According to PSS, concepts develop in the brain as follows. Much research has shown that categories have statistically correlated features (e.g., *wheels*, *steering wheel*, and *engine* for *cars;* McRae, de Sa, & Siedenberg, 1997). Thus, encountering different instances of the same category should activate similar neural patterns in feature systems (cf., Farah & McClelland, 1991; Cree & McRae, 2003). Furthermore, similar populations of conjunctive neurons in the brain's association areas—tuned to these particular conjunctions of features—should tend to capture these similar patterns (Damasio, 1989; Simmons & Barsalou, 2003). Across experiences of a category's instances, this population of conjunctive neurons integrates the modality-specific features of a category, establishing a distributed multi-modal representation of it.

PSS refers to these distributed multi-modal representations as *simulators* (Barsalou, 1999, 2003a, 2005). Conceptually, a simulator functions as a type: It integrates the multimodal content of a category across instances, and provides the ability to interpret later individuals as tokens of the type. Consider the simulator for the category of *cars*. Across learning, visual information about how cars look becomes integrated in the simulator, along with auditory information about how they sound, somatosensory information about how they feel, motor programs for interacting with them, emotional responses to experiencing them, etc. The result is a distributed system throughout the brain's feature and association areas that accumulates modal representations of the category.

In principle, an indefinitely large number of simulators can develop in memory for all forms of knowledge, including objects, properties, settings, events, actions, interoceptions, and so forth. Specifically, a simulator develops for any component of experience that attention selects repeatedly. When attention focuses repeatedly on a type of object in experience, such as *cars*, a simulator develops for it. Analogously, if attention focuses on a type of action (*driving*) or on a type of interception (*fear*), simulators develop to represent it as well. Such flexibility is consistent with Schyns, Goldstone, and Thibaut's (1998) proposal that the cognitive system acquires new properties as they become relevant for categorization. Because selective attention is flexible and open-ended, a simulator develops for any component of experience that attention selects repeatedly.

Once a simulator becomes established for a category, it reenacts small subsets of its content as specific *simulations*. All the content in a simulator never becomes active at once. Instead only a small subset becomes active to represent the category on a particular occasion (cf. Barsalou, 1987, 1989, 1993). For example, the *car* simulator might simulate a jeep on one occasion, whereas on others it might simulate a sedan or a sports car. Because all the experienced content for cars resides implicitly in the *car* simulator, many different subsets can be reenacted as simulations.

The presence of simulators in the brain makes the implementation of symbolic operations possible. Indeed, symbolic operations follow naturally from the presence of simulators. Because simulators are roughly equivalent to concepts, the symbolic functions made possible by concepts are also made possible by simulators. The next three sub-sections illustrate how simulators enable three classic symbolic functions: predication, conceptual combination, and the representation of abstract concepts. For further details, see Barsalou (1999, 2003a, 2005).

3. Implementing the symbolic function of predication in PSS. To implement predication, an intelligent system must first distinguish types from tokens. In PSS, simulators implement types, because they aggregate multi-modal information across category members (e.g., for *cars*). Conversely, simulations implement tokens, because they represent category members (e.g., individual cars). Thus, the simulator-simulation distinction in PSS naturally implements the type-token distinction essential for predication.

This distinction further allows PSS to explain a wide variety of phenomena related to predication, including type-token predication, true vs. false propositions, and interpretive spin. Type-token predication results from binding simulators to simulations (or perceptions). For example, binding the *car* simulator to a simulated (or perceived) car produces the predication that the individual is an instance of the *car* category. These type-token bindings essentially implement propositions, where binding the simulator to the individual represents a claim about the individual, namely, that the individual is a *car*. Notably, such propositions can be false, as when the *car* simulator is applied mistakenly to a small truck. Furthermore, the potential predications of an individual are infinite, thereby producing interpretative spin. Because indefinitely many simulators (and combinations of simulators) could be used to interpret an individual, indefinitely many interpretations are possible. For example, an individual car could be interpreted as a *car*, *vehicle*, *artifact, sedan, junked sedan,* etc. Thus, the simulator-simulation distinction allows PSS to implement classic symbolic functions related to predication.

4. Implementing conceptual combination in PSS. To see how PSS implements conceptual combination, first consider a simulator for the spatial relation, *above*. An *above* simulator could result from having pairs of objects pointed out in perception where the focal object always has a higher vertical position than the other object (e.g., a helicopter above a building). As each *above* relation is pointed out, selective attention focuses on the spatial regions containing the two objects, filters out the objects, and captures modality-specific information about the shapes and sizes of the regions, the vertical distance between them, their horizontal offset, etc. (the parietal lobe would be one obvious location where the *above* simulator might be represented in the brain). Over time, the *above* simulator captures many such pairs of regions and the spatial relations between them. On later occasions, this simulator can produce a wide variety of *above* simulations, each containing a pair of spatial regions not containing objects. An *above* simulation could represent two round regions of equal size, nearly touching vertically, with no horizontal offset; it could represent two rectangular regions of different size, distant vertically, with a large horizontal offset; etc.

The *above* simulator lends itself to producing conceptual combinations. Imagine that this simulator produces a particular *above* simulation. Next imagine that the *helicopter* simulator runs a simulation in the upper region of this *above* simulation, and that the *building* simulator runs a simulation in the lower region. The resulting simulation represents a helicopter above a building, thereby implementing a conceptual combination that expresses the proposition *ABOVE (helicopter, building)* implicitly. Infinitely many other conceptual combinations can be implemented by simulating different kinds of objects or events in the regions of the *above* simulation, thereby expressing related propositions, such as *ABOVE (jet, cloud), ABOVE (lamp, table),* etc. In general, this is how PSS implements conceptual combination. Because simulators represent components of situations and relations between components, their simulations can be combined into complex, multi-component simulations. Much like an object-oriented drawing program, PSS extracts components of situations so that it can later combine them to represent either previously experienced situations or novel ones.

5. Representing abstract concepts in PSS. Relatively little is known about abstract concepts (e.g., *truth*, *thought*), given that most research has addressed concrete concepts (e.g., *car*, *bird*). Abstract concepts, however, are extremely interesting. They are likely to provide deep insights into the nature of human cognition, and to help produce increasingly sophisticated forms of intelligent computation.

Recent theory suggests that one central function of abstract concepts is to represent interoceptive states (e.g., Barsalou, 1999). In an exploratory study, more content about interoceptive states was observed in abstract concepts than in concrete concepts (Barsalou &

Wiemer-Hastings, 2005). In another exploratory study, the abstractness of a concept increased with the amount of interoceptive content that it contained (Wiemer-Hastings, Krug, & Xu, 2001). These studies further found that abstract concepts typically relate interoceptive states to situations and events. For example, *intend* relates interoceptive states about goals to events in the world that follow from them causally (intending to ask someone for information, which then leads to asking the question and receiving an answer).

Because abstract concepts focus on interoceptive states to a large extent, they provide a window on meta-cognition. Similar to how people perceive the external world through vision and audition, people perceive their internal worlds through interoception. During interoception, people perceive motivations, affective states, cognitive states, and cognitive operations. Clearly, only a small subset of brain activity is perceived interoceptively, but this subset supports impressive understanding and control of internal mechanisms.

According to PSS, simulators develop to represent the internal world, just as they develop to represent the external world. As people perceive the internal world, they focus attention on salient aspects of it repeatedly, such that simulators develop for these aspects. Thus, simulators develop for meta-cognitive states, such as *image* and *belief*, cognitive operations such as *retrieve* and *compare*, affective states such as *happiness* and *fear*, motivational states such as *hunger* and *ambition*. Once these simulators exist, they support symbolic operations in the meta-cognitive domain. Simulators become bound to regions of meta-cognitive activity, thereby producing type-token propositions. These categorizations then license associated inferences, and support symbolic analysis of meta-cognitive activity. Predications about meta-cognitive activity result from mapping relevant simulators into regions of it. Finally, novel conceptualizations about how to organize meta-cognitive processing to achieve goals result from combining relevant simulators (i.e., conceptual combination).

6. Summary. These conjectures about abstract concepts and their central role in representing meta-cognition contrast significantly with other views. All other accounts have assumed that abstract concepts are either represented with amodal symbols or with language. Furthermore, no previous account has proposed that abstract concepts are central to meta-cognition.

The research performed under this DARPA contract focused on whether the three symbolic operations just described—predication, conceptual combination, and the representation of abstract concepts—are grounded in simulation as PSS predicts. The next three sections present empirical results that support this account. Besides offering empirical support for the predictions of PSS, these experiments also offer methodological innovations for performing research on grounded cognition.

II. Evidence for Simulation in Predication

Two lines of research were developed under this DARPA contract to assess whether the symbolic operation of predication is grounded in simulation. One line of research used a behavioral paradigm, and the other used fMRI. Both paradigms are novel, not having been used by other researchers or ourselves previously. Both paradigms offer much potential for studying the fundamental process of predication and for assessing theoretical accounts of it. Each paradigm, along with results obtained with it to date, is addressed in turn.

A. Behavioral Evidence for Simulation in Property Predication

As described earlier, predication results from binding a concept to an individual (e.g., binding the concept of *windshield* to a particular car, thereby predicating *windshield* of the car). As also described earlier, predication is generally assumed to result from binding an amodal symbol for a concept to an individual, as in *windshield* (*X*). Notably, such symbols are assumed to abstract over the details of their referents, thereby standing for all of them. Thus, the amodal symbol for *windshield* abstracts over the particular details of different windshields.

Conversely, PSS assumes that the concept for *windshield* is a simulator that has integrated perceptual detail about windshields across many instances. As a result, predicating *windshield* of a referent should activate this perceptual information during the predication process. Furthermore, if windshields of a particular type have been perceived more often than others (e.g., tinted windshields), then this frequent perceptual information should become active during predication and be extended to predicated individuals. Unlike standard amodal views of predication that are insensitive to minor perceptual variation in the instances of a concept, the PSS view assumes that predication should be sensitive to such variation. Because the concepts underlying predication contain perceptual detail, perceptual detail should affect the predication process.

The paradigm developed for this project provides a means of testing this prediction. To our knowledge this is a novel paradigm that has not been used before. Furthermore, this paradigm is sufficiently simple that autonomous computational agents could be expected to perform it. If the cognitive systems of these agents were based on the simulation architecture, these agents would show the perceptual bias predicted for these experiments. In general, this paradigm can be used to assess whether concepts acquired from experience contain subtle perceptual detail. If they do, then this supports the PSS account of predication.

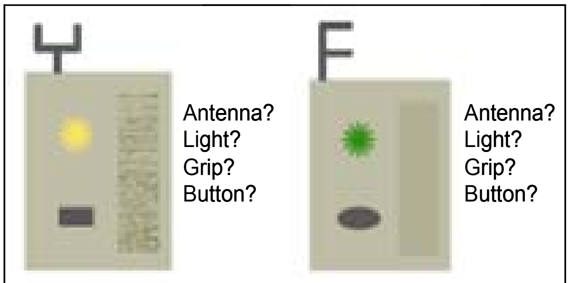


Figure 3

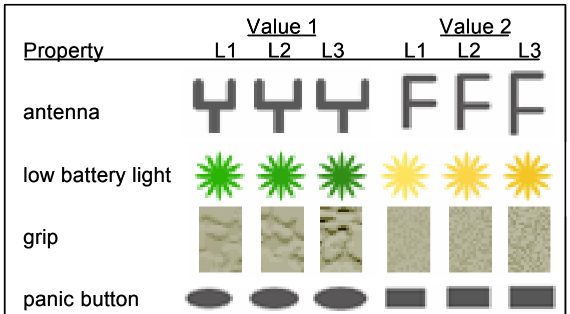


Figure 4

1. The basic paradigm. The experiments in this project typically contain three phases: (1) bias, (2) study, and (3) test. Each phase is described in turn.

Bias phase. In the bias phase, participants are perceptually exposed to a novel type of object that they probably have not experienced before—a spy device—and its properties. Their task is learning to predicate properties correctly of the spy devices. Later phases of the experiment assess whether the predicates learned for these properties contain perceptual information or not. Figure 3 shows two examples of these devices. A cover story motivates the purpose of the device, its components, and their functions.

Four critical properties vary across devices (antenna, battery light, grip, panic button). As Figure 4 illustrates, each property has two values, with each value having three levels. The antenna has two shape values (U vs. F antenna), with the distance between the parallel bars for each varying very slightly across three levels. The low battery light has two color values (green vs. yellow), with each color having three slightly differing levels of hue. The grip—the rectangle along the right edge of the box—has two texture values (craqueled vs. grainy), with each having three levels of coarseness. The panic button has two shape values (oval vs. rectangle), each having three levels of area.

Perceptual bias is implemented for each of the eight property values in Figure 1. This bias will be central to assessing the PSS account of predication. It is important to note that different participants in these experiments receive different biases, such that bias is carefully controlled.

Table 1 illustrates how bias for each value of a property is implemented. In Version 1 of the materials, a distribution of levels is used that biases U antennas towards Level 1. As can be seen, Level 1 occurs 18 times, whereas Levels 2 and 3 do not occur at all. Across the 18 devices that have a U antenna, participants should develop a perceptual bias that associates U antenna with Level 1 (assuming that the PSS account of predication is correct). Also in Version 1, yellow battery lights are biased towards Level 1, whereas craqueled grips and oval panic buttons are biased towards Level 3. The other four property values in Version 1 receive the opposite bias, namely, F antennae and green battery lights are biased towards Level 3, whereas grainy grips and rectangular panic buttons are biased towards Level 1. For control and counter-balancing purposes, Version 2 (received by a different group of participants) has the opposite assignments.

Table 2 illustrates a subset of the 36 trials in one randomized presentation sequence for Version 1. As can be seen, each participant studies 36 different spy devices. Across devices, each value of a property is biased towards either Level 1 or Level 3. No correlation between values exists in this experiment.

Version 1 of the Materials

Property + Value	L1	L2	L3	Property + Value	L1	L2	L3
	Frequency of Value Level				Frequency of Value Level		
U antenna	18	0	0	F antenna	0	0	18
yellow battery light	18	0	0	green battery light	0	0	18
craqueled grip	0	0	18	grainy grip	18	0	0
oval panic button	0	0	18	rectangular panic button	18	0	0

Version 2 of the Materials

Property + Value	L1	L2	L3	Property + Value	L1	L2	L3
	Frequency of Value Level				Frequency of Value Level		
U antenna	0	0	18	F antenna	18	0	0
yellow battery light	0	0	18	green battery light	18	0	0
craqueled grip	18	0	0	grainy grip	0	0	18
oval panic button	18	0	0	rectangular panic button	0	0	18

Table 1

On each trial, a device is shown for 3 sec. Then, while the device remains on the screen, each of its four properties is queried. As Figure 3 illustrates, a sequential series of queries appears vertically on the screen for 3 sec each: Antenna? Light? Grip? Button? For each query, the participant states the value of the property. In response to Antenna?, for example, a participant states verbally that it is either a U or F antenna. The purpose of these queries is to create an association between each verbal label and the biased level of its value. Across the 18 trials when a U antenna is shown, participants receiving Version 1 should associate the verbal label, "U antenna," with Level 1. Later, according to PSS, these biases should be triggered when U antenna is predicated of new spy devices.

Trial	Antenna	Battery Light	Grip	Panic Button
1	F-3	G-3	G-1	R-1
2	F-3	Y-1	G-1	R-1
3	U-1	G-1	C-3	O-3
4	F-3	Y-3	G-1	R-1
5	F-3	Y-1	C-3	O-3
6	U-1	G-3	G-1	R-1
7	U-1	Y-1	G-1	R-1
8	F-3	Y-1	C-3	O-3
.	.	.	.	.
36	F-3	Y-1	C-3	R-3

Table 2

Study phase. Participants are told that the next phase of the experiment involves learning to identify the property values of devices belonging *to particular spies*. Nothing is said about a subsequent memory test. Participants study two devices, one for each of two different spies (CIA-99 and KGB-50). Across the two devices, each of the two values for the four properties is presented once. For example, CIA-99's device might have an F antenna, a yellow light, a craqueled grip, and a rectangular button, whereas KGB-50's device might have a U antenna, a green light, a grainy grip, and an oval button. Importantly, however, the level for each value is always Level 2, which lies half way between Levels 1 and 3 subjectively. Thus, the values shown for the two devices belonging to CIA-99 and KGB-50 were not seen earlier during the bias phase (although they are similar). Furthermore, the values shown for these devices lie between the biased values for the two different versions of the materials.

Presentation of the two studied devices is the same as in the bias phase. First, each device is shown for 3 sec, and then each of its four critical properties are queried sequentially for 3 sec each. Labeling the four properties of each device in response to these queries implements the symbolic activity of interest: predication.

The key prediction is as follows. Generating the label, "F antenna," for CIA-99's device should activate the biased form of F antennas stored in memory during the bias phase (e.g., L3 for Version 1). If the simulation account is correct, this biased value should be simulated on producing the label, such that it becomes bound to the L2 value in the studied device during predication. As a result, this fusion should later distort memory of the F antenna towards L3. Conversely, when participants receiving Version 2 of the materials generate the label, "F antenna", this should activate a simulation of L1, which becomes bound to the L2 value in the studied device, biasing later memory of it towards L1. Alternatively, if an amodal symbol represents the concept, *F-antenna,* it should not be affected by these minor perceptual variations in perceptual bias. Traditional accounts of assume that amodal symbols abstract over the kind of perceptual detail varied here, such that predicating *F-antenna* should activate the same amodal symbol in both bias conditions.

Test phase. Following the study phase, participants perform a buffer task for 10 minutes (watch a segment of a spy movie and answer questions about it). Afterwards, participants receive two tests of their memory for the two studied spy devices: (1) object recognition, then (2) property recognition.

On the object recognition test, participants perform two forced-choice trials, one each for the devices belonging CIA-99 and KGB-50. The left side of Figure 5 illustrates these trials. Participants see three devices and are asked which belonged to a particular spy. Consider the trial for CIA-99's device. As the left panel of Figure 5 illustrates, the three devices in the choice set all have the same values (i.e., F antenna, yellow light, craqueled grip, rectangular button). One of the three devices is the device seen during the study phase for CIA-99 (choice B). Again, all four values for this particular device have level L2.

A second device in the choice set (choice C in Figure 5) contains the biased level for each property value from the bias phase. Thus, for Version 1, this device contains L1 for yellow light and rectangular button and L3 for F antenna and craqueled grip (see Table 1).

The third device in the choice set (choice A in Figure 5) contains the opposite of the biased level for each property value, lying on the other side of L2. Thus, for Version 1, this third choice contains L3 for yellow light and rectangular button and L1 for F antenna and craqueled grip (see Table 1).

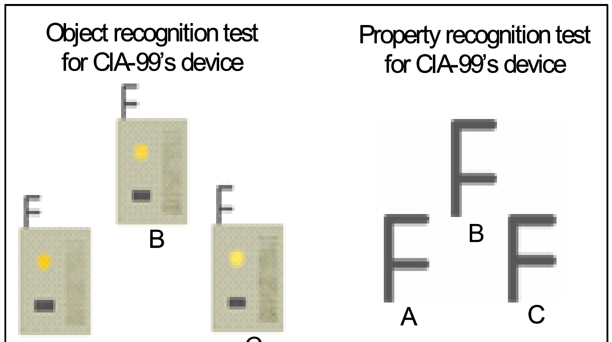


Figure 5

Participants' task is, first, to select the device that they think belonged to CIA-99. Once they make a choice, they then select the device from the two remaining that they think is next most likely to have belonged to this spy. These two choices thus rank the three test stimuli 1, 2, and 3. After completing the first forced choice, participants perform an analogous test for KGB-50's device.

If participants use amodal symbols to predicate the property values of CIA-99's device, they should randomly choose a device from the choice set. Because amodal symbols abstract over minor perceptual details, minor variations in property value level should not enter into processing, such that no bias occurs. If, however, participants use simulators to perform predication, they should choose the device that contains the four biased values. According to this account, when participants predicate property values of CIA-99's device during the study phase, the predicates that participants use project biased perceptual information onto the device's actual properties. As a result, biased perceptual information in the predicates becomes fused with the perceived property values of the device. Later at test, memory distortion occurs, when both the actual and predicated values are retrieved.

Additional tests similarly assess memory for the four individual values of each spy's device (as opposed to the entire device). As the right panel of Figure 5 illustrates, participants received the three levels of each property, and had to rank them according to their likelihood of belonging to a particular spy device. Analogous to the full object tests, one choice was the L2 value presented in the spy's actually device, a second choice was the biased level seen during the bias phase, and the third choice was the non-biased value not seen during the study phase. Again, the prediction is that if predication relies on simulators, then participants should tend to believe that the biased values were presented for the spy devices seen during the study phase, when in fact they were not.

All tests are fully counter-balanced. In the object recognition test, the order of the two objects is counter-balanced, as are the spatial positions of the three choices in each choice set. In the property recognition test, these factors are again counter-balanced, as is the order in which individual properties are tested.

Interpretations of the predicted result. If the predicted bias effect emerges, it suggests that perceptual simulations underlie the conceptual content of predicated properties. Alternatively, however, it could be argued that amodal symbols represent these properties, accompanied by perceptual memories. Notably, however, amodal theories do not predict such bias effects a priori (Barsalou, 1999). Instead, the spirit of these theories is that a discrete amodal symbol represents each property value (e.g., *craqueled*), abstracting over perceptual details, such as slightly varying degrees of coarseness. Amodal theories most naturally predict that, during the bias phase, participants establish an amodal symbol for each property value of the spy devices, with a single discrete symbol standing for all its different levels.

Later, when a craqueled grip is labeled during the study phase, the label activates the respective amodal symbol for *craqueled* to represent the property in a memory of the studied device. If so, there should be no bias, given that a discrete symbol, which stands for all the different levels of *craqueled*, represents this property in memory—no information about the bias is included. Amodal theories do not generally predict that perceptual memories become active with symbols, which then produce bias.

In contrast, the PSS account explains these bias effects naturally and parsimoniously, using the construct of a simulator whose biased simulations of perceptual information become bound to regions of perceptions and memories during predication. Amodal symbols are not needed to play any functions that simulators cannot already perform.

2. Establishing the basic bias phenomenon. A first experiment was performed as just described above using 24 participants. The results on both the object recognition test and property recognition test showed strong effects of perceptual bias.

First consider the results for the object recognition test. A weighted contrast was used to test whether the biased property values distorted memory for the two studied objects. Specifically, the contrast assessed whether participants' rankings of the three test objects were correlated with bias from the bias phase. In these contrasts, the device having the four biased values on the object recognition test was assigned the value of 1. The studied (neutral) device was assigned 0. The device having the opposite of the distorted values was assigned –1 (because it had less bias than the neutral device). These weights were then multiplied with participants' ranks for the choices. The device that a participant selected as most likely to have been studied was weighted 1, the device judged next most likely was weighted 0, and the device judged least likely was weighted -1.

If biased property values distorted participants' memories of the objects, then the bias ranks and participants' ranks should be correlated, such that the weighted contrast exhibits values significantly greater than 0. For example, when participants select the biased object first, and the studied object second, the contrast is (1 x 1) + (0 x 0) + (-1 x -1) = 2. Similarly, when participants select the biased object first and its opposite second, the contrast is (1 x 1) + (-1 x 0) + (0 x-1) = 1. Alternatively, if predication uses discrete amodal symbols—not biased perceptual values—this contrast should not differ significantly from 0. There should be no tendency for bias to correlate with the rankings. Thus, the contrast used to assess the hypothesis of interest ranged from 2 (complete bias) to 0 (no bias) to -2 (the opposite of the predicted bias).

As Figure 6 illustrates, the contrast averaged 1.20 in the first experiment, being significantly greater than 0 ($t(23) = 5.33$, SE=.22, $p<.001$). This finding indicates that participants had highly biased memories of the devices in the study phase. They did not remember the study stimuli as they had been presented. This finding supports the a priori prediction of PSS that simulations constitute the conceptual content of the predications made during the study phase.

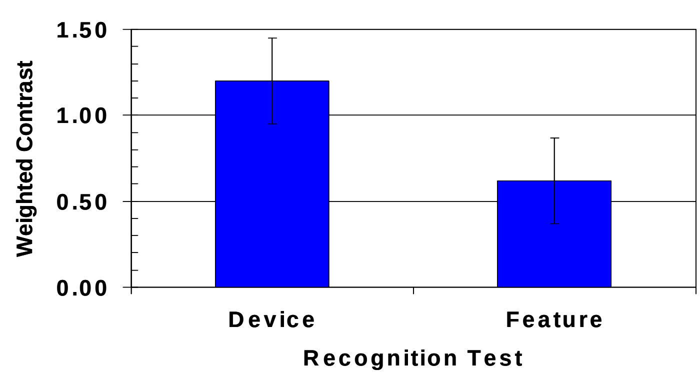


Figure 6. Results for the forced-choice recognition tests of devices and features.

Next consider the results for the feature recognition test. The same contrast was computed for each set of three choices that a participant made across the eight tested properties. Figure 6 shows that the average of these contrasts across properties and participants, .62, was again significant ($t(23) = 2.21$, SE=.28, $p<.025$). This finding further corroborates the PSS prediction that property predicates applied during the study phase contained perceptual information.

Interestingly, the amount of bias for entire spy devices (1.20) was twice the bias observed for the individual features of spy devices (.62). This suggests that when four features are predicated together, bias aggregates across the four features to produce stronger bias overall than occurs for one feature alone.

In summary, this first experiment indicates that the concepts used for predicating properties contain perceptual information. Consistent with PSS, the acquisition of new property concepts is biased toward the perceptual information of instances encountered during learning. Later predicating the property of a slightly different property value produces distortion towards the earlier biased value. This finding supports the proposal that classic symbolic operations, such as predication, are implemented by modality-specific mechanisms, not by amodal symbols.

3. Assessing the role of language. One possibility is that language plays an important role in the bias effect. As properties of the spy devices are acquired during the bias phase, not only is perceptual information acquired for them, so are linguistic labels (e.g., “craqueled grip”). Perhaps the perceptual interference observed in the first experiment resulted from participants applying these labels during the study phase, which in turn triggered simulators that represent the properties conceptually. Once the simulators became active, they produced simulations, which then biased memory of the studied neutral properties.

Alternatively, language may not be necessary for triggering these biasing effects. Instead, mere perception of a property during the study of a particular spy’s device may be sufficient to activate a simulator that biases memory of it. In other words, labeling the property linguistically may not be necessary to activate the simulators that later produce bias. If purely perceptual triggering is sufficient, this would have significant implications for how people process perceptual experience independently of language. It would suggest that merely perceiving the world (without describing it linguistically) activates simulators that predicate properties of perceptual experience, which in turn distort it. This would also have implications for the design of cognitive agents, who could be built to implement this same kind of perceptual distortion. Although such distortion might appear undesirable, it could play useful roles, such creating coherent perceptual experiences, distinguishing familiar situations from unfamiliar situations, etc.

Method. The same basic paradigm used in the first experiment was also used here. Indeed, one condition offered a near replication. However, four different groups of participants were run in a two-by-two design created by crossing the following two manipulations orthogonally. First, verbal vs. visual encoding was manipulated during the bias phase to see if the presence of language during predicate learning is necessary for perceptual bias later in the test phase.

Verbal encoding used nearly the same learning procedure during the bias phase as in the first experiment. On each trial, a spy device was shown for 3 sec. Each of the device’s four properties was then probed in a random sequence as follows. First, the name of the property (e.g., “Grip?”) appeared at the top center of the screen, with the device below, for 3 sec. The device then disappeared while the property name remained on the screen with the names of the two possible values below, one on each side of the screen, determined randomly (e.g., “Craqueled” on the left, “Grainy” on the right). The participant then had 3 sec to press a response key on the left or right to indicate which value (on the left or right) had appeared for the previous device. After the 3 sec choice period, the incorrect answer disappeared and the correct answer remained for 1 sec, followed by a 1 sec blank period. Each remaining feature was tested similarly until all four features had been tested. When finished, participants received the next device and predicated its properties similarly. Like the encoding task in the first experiment, this task should create strong associations between the linguistic labels of property values and the corresponding perceptual bias.

Visual encoding used a very different method of learning properties during the bias phase. During the entire trial for processing a device, participants performed articulatory suppression to prevent (or at least minimize) verbal processing. At the start of each trial, “Begin” was shown for 3 sec and participants began uttering “the the the…” at continuous 1 sec rate. As much work has shown, this continuous articulation should prevent the verbalization of features during the trial, or at least decrease verbalization to a much lower level than in the verbal encoding task, where verbalization was encouraged. After the initial 3 sec articulation period, a spy device was shown for 3 sec. The word “Study” then appeared at the top of the screen with the device below for 3 sec, while participants studied the device for the subsequent imagery period. After the study period ended, "Close eyes and image" appeared that the top of the screen, and the device disappeared. During this 3 sec period, the participant was asked to mentally image how the device had looked (participants were told that this would help them learn about the appearance of the devices). After the mental imagery period ended, a tone played for 1 sec, the participant opened his or her eyes, and a blank screen appeared for 1 sec. This study-image cycle continued three additional times, so that the total presentation time was the same as in the verbal encoding condition. After the fourth repetition of this sequence, the trial ended, the participant stopped saying “the,” and he or she waited until the next trial began. Thus, the instructional set in this condition oriented participants away from verbalizing the properties and towards studying and imaging the devices visually.

The sequence of device presentations on a given trial was identical in the verbal and visual encoding conditions. The only difference between the two conditions was that, in the verbal encoding condition, words for the properties and their values were presented, and participants had to select the property values shown for the device. Of primary interest was whether participants in the visual encoding condition would still show perceptual bias later in the test phase. If linguistic labeling is necessary to activate the simulators that bias memory during the study phase, then the visual encoding condition should not exhibit bias, or at least much less than the verbal encoding condition.

This same manipulation between verbal and visual encoding was also implemented during the study phase. As participants studied the two spy devices belonging to CIA-99 and KGB-50, they either performed verbal or visual encoding. As Table 3 shows, the manipulations of verbal vs. visual encoding during the bias and study phase were crossed orthogonally between participants to produce four between-participant conditions of 24 participants each.

	Study Phase Encoding	
Bias Phase Encoding	Verbal	Visual
Verbal	24 participants	24 participants
Visual	24 participants	24 participants

Table 3. Design of second behavioral experiment on predication

If verbal encoding is necessary for perceptual interference to occur, then as more verbal encoding is performed, more interference should occur. Specifically, participants in the verbal bias / verbal study condition should show the most bias, followed by participants in the verbal bias / visual study condition and in the visual bias / verbal study condition. Participants in the visual bias / visual study condition should the least bias.

Results. Contrary to the prediction that language is necessary for triggering perceptual bias, equal bias occurred in all four conditions. As Figure 7A illustrates, perceptual bias was significantly greater than 0 in all four conditions. Furthermore, perceptual bias was just as high for the visual bias / visual study condition as for the verbal bias / verbal study condition. The other two mixed conditions exhibited similar levels of bias as well. Indeed, there were no significant differences between any of the four conditions. This pattern indicates that language is *not* necessary for perceptual bias. Even when the use of language is eliminated during the bias and study phases (or at least reduced greatly), perceptual bias still occurs.

This finding indicates that simply perceiving spy devices during the bias phase—without verbally describing them—encodes biased perceptual memories of the devices and their properties. Similarly, perceiving the devices of particular spies during the study phase activates this biased information without these devices being described. Participants appear to acquire predicates for the properties during the bias phase and then apply them to “neutral” devices during the study phase without the use of language (or at least with greatly reduced language).

This pattern suggests that the process of forming and applying simulators can operate independently of language, perhaps through the use of selective attention. As participants in the visual bias condition study the training devices, selective attention focuses on the four properties, extracts perceptual information from them, and stores this information in the simulator established for the property on previous trials. As a result of simply focusing attention on these properties, simulators develop in memory for them. Later, during the study phase, as the neutral devices of particular spies are studied, these simulators become active as a result of focusing attention on a device’s properties, and distort memory of the actual properties studied.

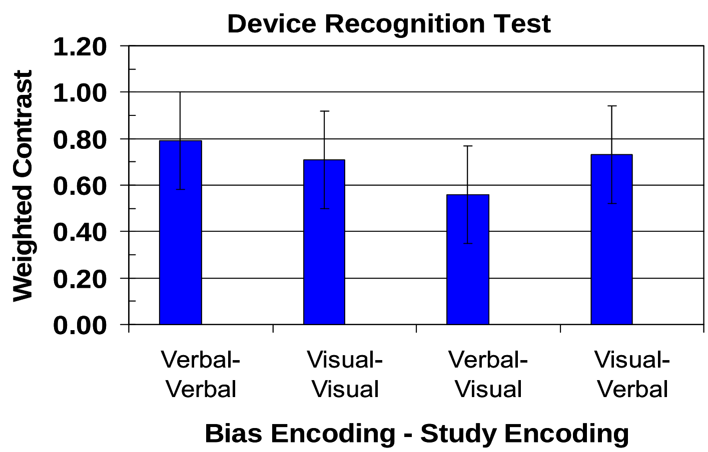


Figure 7A. Results for the forced-choice recognition test of devices.

The construction and application of simulators using attention alone—without language—may be a basic cognitive process. It may operate extensively in pre-linguistic infants and in non-humans. It may also operate frequently in adults on aspects of perceived experience for which words do not exist. Even when words do exist, this basic mechanism may operate prior to, or at least in conjunction with, the use of linguistic labeling mechanisms.

As Figure 7B illustrates, significant bias also occurred on the feature test. Furthermore, the four conditions generally did not differ from each other, although the verbal bias / visual study group did show significantly more bias than the verbal / verbal and visual / visual group for reasons we do not understand. Again, as in the first experiment, bias on the feature recognition test was less than on the device recognition test. As suggested there, we believe that the greater bias on the device test results from the aggregation of bias across four features.

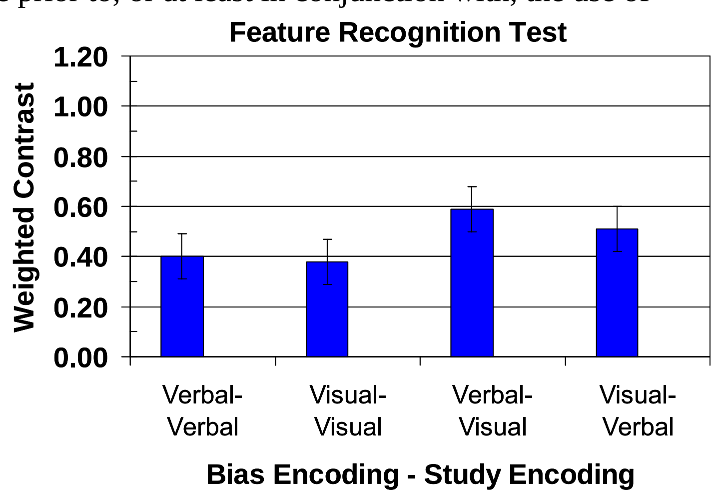


Figure 7B. Results for the forced-choice recognition test of features.

4. Perceptual inference at encoding vs. retrieval. In this next line of research, we addressed when, during processing, perceptual interference occurs. One possibility is that interference arises at encoding. As participants study the critical spy devices, simulators acquired for properties during the bias phase predicate relevant aspects of the device. During this predication process, simulations from the simulators become merged with properties in the devices, distorting them. Alternatively, interference could arise at test. When participants study the three test choices, simulators could activate prototypical values of the properties that interfere with episodic memories of the target device. As a result, participants exhibit bias towards recognition choices that contain biased property values. Finally, simulators could affect both encoding and retrieval, activating interfering information at both times.

Ava Santos' dissertation—completed and being defended in early August, 2007—included an experiment that addressed this issue. The same spy device paradigm used in previous experiments was used here. Figure 8 illustrates the design of this study. Participants first studied the devices of four spies in the "early" study phase, prior to receiving any form of property bias. As a result, these devices could not be encoded in a biased manner. Second, participants acquired biased property information as in the previous experiments. Third, participants studied another four devices in the "late" study phase. Because biasing information had just been acquired, it could have been used to distort the late studied devices as they were encoded. Fourth, and finally, participants' memory of the eight studied devices was tested (four devices from the early study phase, and four from the late study phase; the test phase is not shown in Figure 8). Because biasing information had been acquired prior to this test, it was potentially available to distort memory for all eight devices.

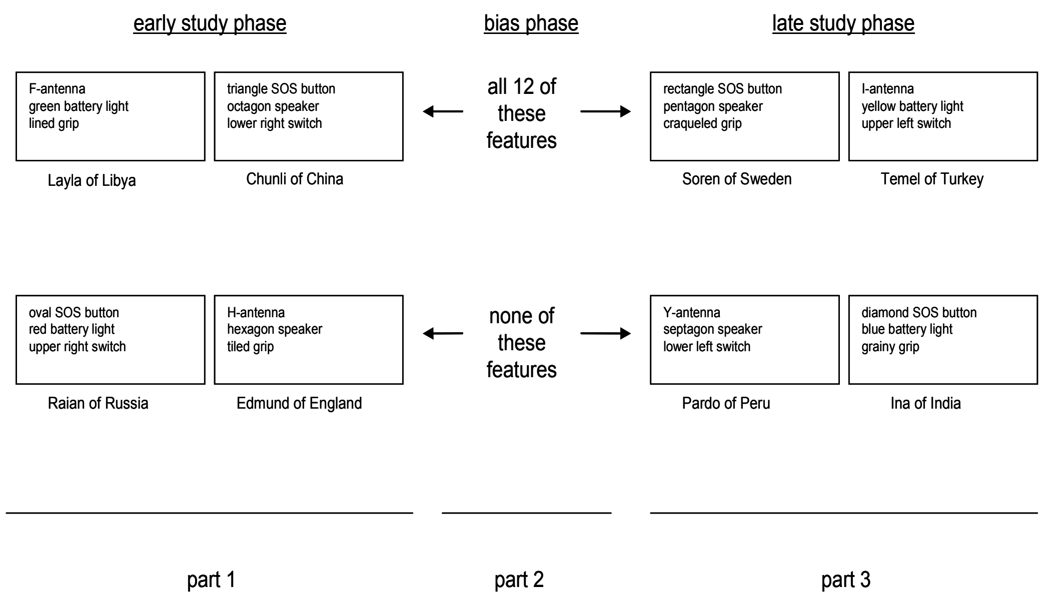


Figure 8. Illustration of the design in an experiment that assessed whether perceptual interference occurs at encoding, retrieval, or both.

If perceptual interference occurs at encoding, then interference should only occur for the devices in the late study phase. Because biasing information was not available while encoding devices in the early study phase, they should not experience interference. Conversely, if perceptual interference occurs at retrieval, then interference should occur for the devices in both the early and late study phases. Because biasing information was available for devices in both phases, memory for all eight devices should be distorted. Finally, if bias occurs at both encoding and retrieval, then memory for devices in both phases should be distorted, but greater distortion should occur for devices in the late study phase. Because devices from the late study phase experience interference at both encoding and retrieval, they should suffer more distortion than devices from the early study phase that experience interference only at retrieval.

Control conditions were included to assess whether memory varied over time from the early to late study phases. If it did, then this could complicate interpreting the amount of bias in each study phase. Including control conditions allowed us to assess the amount of bias in each phase relative to memory accuracy at that time. As Figure 8 illustrates, the eight devices in the control conditions (four early and four late) contained properties that were never biased during the experiment. Thus, memory for these control devices should primarily reflect basic memory processes unaffected by distortion created within the experiment. When assessing the key predictions about encoding and retrieval, bias in the critical conditions was measured relative to bias in the control conditions. If bias occurs at encoding, then bias in the late study condition should be significantly higher than in the late control condition, but bias in the early study condition should not differ from bias in the early control. If bias occurs at retrieval, then bias in the both the early and late study conditions should be significantly higher than in the early and late control conditions, respectively. If bias occurs at both encoding and retrieval, then bias in both the early and late study conditions should be significantly higher than in their respective control conditions, but the amount of bias relative to control should be significantly greater for late study than for early study.

Figure 9 shows the results of this experiment. As can be seen, the results indicate that bias only occurred at retrieval not at encoding. Specifically, significant bias occurred during both the early and late study phases, relative to control, consistent with the retrieval hypothesis. However, the amount of bias in the late study phase was not greater than the bias in the early study phase (again, relative to the respective controls). Thus, encoding did not produce additional bias above and beyond the bias at retrieval.

These results are interesting and informative. Theoretically, they indicate where the causal effects of interference occur. Besides having implications for our experiments, they shed new light, and in cases new explanations, on related findings in the literature. At a more practical level, these results are useful in designing applications that minimize interference. We now know that we need to build in protection against bias from interfering memories when memory is tested. It's also quite interesting that interference does not appear to be occurring at encoding (but see the next experiment). This was surprising to us and not expected. Further research should explore why

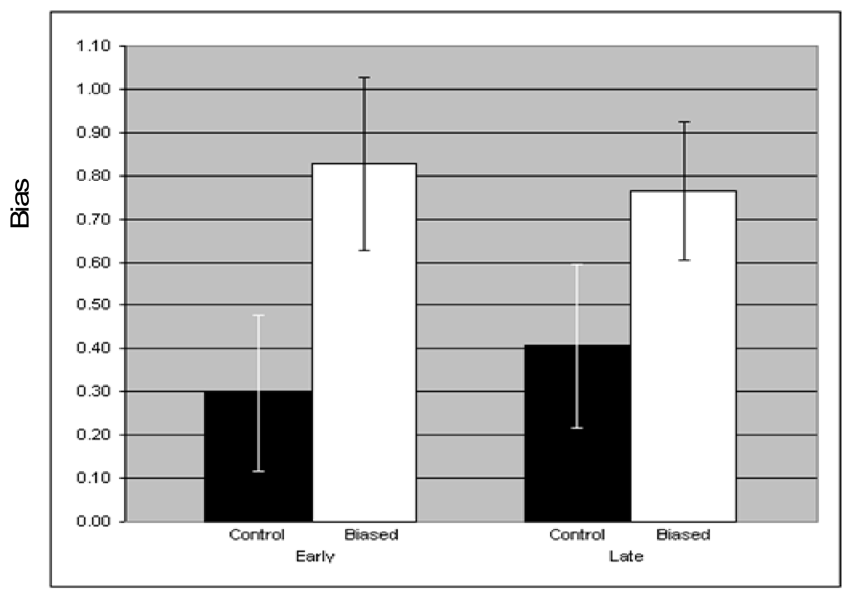


Figure 9. Bias at recognition in the experiment that assessed whether perceptual interference occurs at encoding, retrieval, or both.

interference at encoding does not appear occur in this particular form of the paradigm. One possibility is that the explicitness of a memory test is required for the interfering information to have effect.

5. Bottom-up and top-down modulation of perceptual interference. The second experiment in Ava Santos' dissertation addressed whether perceptual interference is modulated by bottom-up and top-down factors. This experiment followed the standard paradigm used in the first two experiments. Specifically, participants first received bias, then studied critical spy devices belonging to particular spies, and finally had their memory of the critical spy devices tested. Unlike the previous experiment, but like all other previous experiments, there was no early study phase, only the standard late one.

The bottom-up factor assessed was the amount of presentation for a studied device. The longer that a studied device is presented physically during the study phase, the more veridical information that participants can extract from it. As a result, bias should decrease relative to a condition in which presentation time was shorter. In other words, as the perceptual information extracted bottom-up from a studied device increases, the relative ratio of veridical information to biased information increases, such that memory should be increasingly accurate. To implement this manipulation, some studied devices were presented four times, whereas other studied devices were presented just once. In the last three presentation phases, no predication was performed of the devices. They were simply studied visually, so that no additional bias from predication would accrue along with the accumulating bottom-up information.

The top-down factor assessed was the amount of time that participants predicated properties of a device. The more predication that occurs, the more that bias should occur. As a result, bias should increase relative to a condition that has less predication. In other words, as the bias generated top-down from biased property simulators increases, the relative ratio of biased information to veridical information increases, such that memory should be increasingly distorted. To implement this manipulation, the properties of some studied devices were predicated four times, whereas the properties of other studied devices were predicated just once. In the last three predication phases, the device was not presented visually, thereby preventing additional bottom-up information from accruing as well.

Figure 10 presents the results of this experiment. The bars on the left labeled "visual" show the results when the veridical stimulus information was presented once (baseline) vs. multiple times (multi). As can be seen, increasing the amount of bottom-up information available decreased bias. Less bias occurred in the multiple presentation condition than in the single presentation condition. The bars on the right labeled "label" show the results when biased predicates are applied once (baseline) vs. multiple times (multi). As can be seen, increasing the amount of top-down bias from memory increased bias. More bias occurred when participants used predicates to label the devices multiple times than when they only labeled them once.

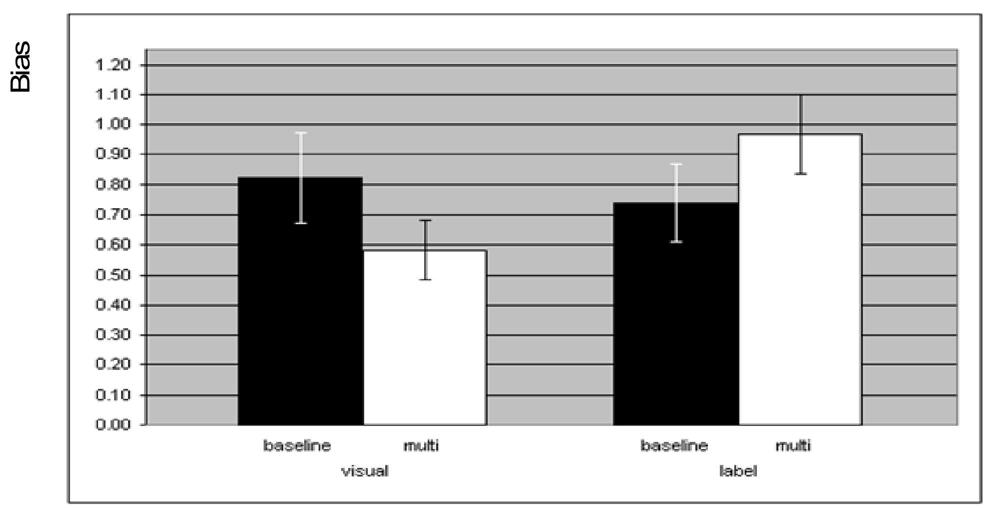


Figure 10. Bias at recognition in the experiment that assessed modulating influences of bottom-up presentation time and top-down predication time.

This experiment demonstrates that bottom-up and top-down factors modulate perceptual interference that arises during predication. Because predication is affected by these perceptual manipulations, perceptual information is implicated in the content and use of predicates, consistent with PSS.

Again, this experiment also has applied implications. To avoid memory distortion, the amount of original study should be as extended as much as possible, thereby encoding maximal amounts of veridical information into memory. Furthermore, minimizing subsequent predication about the memory will also reduce distortion. As this experiment illustrates, the more predication applied to a memory, the more opportunities there are for bias to occur.

Interestingly, the bottom-up and top-down manipulations in this experiment were both encoding manipulations, given that they occurred when the critical devices were studied, not when they were tested. This indicates, contrary to the previous experiment, that encoding effects do occur, at least when certain manipulations are performed at encoding, such as extended study and extended predication. One possibility is that encoding effects in the previous experiment were masked by retrieval effects that were large enough to create the maximal interfering effects possible, such that encoding effects could not be detected. Further research and theory are required to resolve these issues.

6. Perceptual interference for faces. Another graduate student, Shlomit Finkelstein, and I extended the perceptual interference paradigm to faces. This is important for a number of reasons. First, much work on perceptual interference (verbal overshadowing research in particular) has focused on faces, such that it would be useful to replicate our effects in this domain. Second, it is important to replicate our effects in other domains, to ensure that they do not simply arise from idiosyncratic factors in one domain (in this case, the domain of artificial spy devices). Third, there is tremendous interest in face processing and face memory in multiple research communities. By demonstrating perceptual interference for faces, we can significantly increase the visibility of this phenomenon. Fourth, face processing is of considerable interest for both social and applied reasons, and our results have significant implications for these areas.

Our experiment with faces used the same design as the first experiment in this series. Participants first received biasing information about the properties while learning to predicate property values of faces (particular values of eyes, noses, cheeks, chins, hair, ears, etc.). Participants then studied the faces of several named individuals for a later memory test. Finally, participants' memory of the named individuals was tested. Of interest was whether memory of the named individuals' faces was biased towards the biased facial properties acquired during the bias phase of the experiment. So that we could subtly create slightly different facial properties and control them parametrically, we used Poser, a software package well-suited to this purpose. Figure 11 presents examples of our facial stimuli constructed with this software. We plan to use this carefully constructed stimulus set in future experiments.

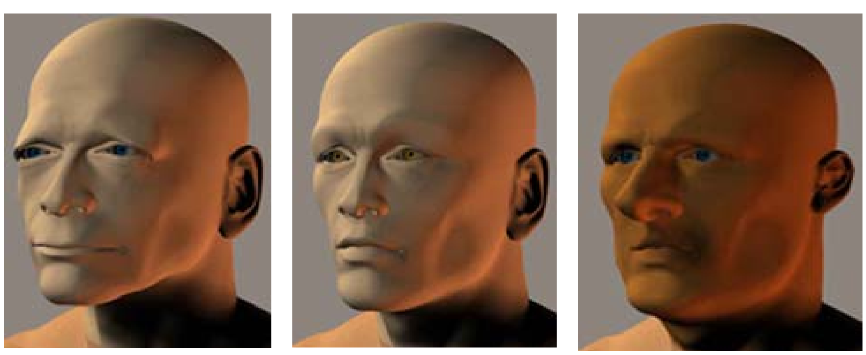


Figure 11. Examples of the face stimuli used in the verbal interference experiment on faces.

Analogous to the first experiment in this series (on which this last experiment was modeled), significant distortion occurred at test. Specifically, the amount of bias, measured by the bias contrast, was .73, significantly greater than 0. Thus, bias not only occurs in the spy device paradigm, it also occurs for faces, indicating that perceptual interference during predication is a general phenomenon.

7. Further experiments. This paradigm has much potential to explore a variety of issues concerning the role of perceptual simulation in the representation and processing of concepts. Further experiments could assess: (1) whether language triggers the presence of perceptual bias after biasing information has become relatively inaccessible in memory; (2) whether abstract concepts, such as relations, exhibit perceptual bias. Many further applications of this paradigm are possible to explore many other issues.

B. fMRI Evidence for Simulation in Category Predication

We just saw behavioral evidence that the process of predication utilizes perceptual information. Consistent with PSS, learning a new property and experiencing perceptual information for it establishes a concept in memory that contains this information. This concept does not appear to be an amodal symbol that transcends perceptual information. Once a perceptually-grounded concept exists in memory, it can then be used for predication, interpreting features of later objects as instances of it.

This next line of work addresses several related issues at the neural level. First, when a person learns a new concept, where is its conceptual content stored in the brain? If PSS is correct, it should be stored in perceptual areas. Second, when the word for the concept is read (but no instance of the concept actually shown), does the word activate perceptual areas of the brain to represent its meaning, as PSS predicts?

The methods for addressing these questions have changed considerably since those described in the original proposal. After encountering unanticipated methodological problems, we went through a long process of evaluating various paradigms. We eventually settled on a paradigm that is superior to the original in many ways. To our knowledge, nothing like this paradigm has ever been implemented before. In our opinion, it offers a useful new tool for performing neuroimaging studies of complex learning tasks, like ours. We also ran this experiment on the Emory scanner after it was upgraded to perform high resolution imaging. As a result, we were able to examine brain areas of interest in much more detail than would have been possible previously. Finally, analysis of the data from this experiment when through several phases until we finally converged on an approach that was rigorous, avoided various pit falls, and that tested our hypotheses appropriately.

1. Experiment overview and materials. Similar to the behavioral experiments in the first project, training phases preceded the critical test phase. In the experiment here, all of the training phases were performed outside the scanner, and then the critical test phase was performed in the scanner. During the training phases, participants learned the three artificial categories illustrated in Panel A of Figure 12, which shows examples of members of each category. As can be seen, the members of a category have a common underlying structure, but differ in the details of how the common structure is realized. As can also be seen, the members of a category also vary in their orientation around vertical. The digital art program, Twisted Brush, which has thousands of different virtual “brushes” for creating the same form in different ways, was used to create the stimuli. As readers familiar with Chinese will note, the three categories are Chinese calligraphy characters. Because none of our participants knew Chinese, however, these categories were novel for them. As Figure 12 further illustrates, each category was associated with a nonsense syllable (dax, jid, or wul). Thus, both the categories and their names were unfamiliar to participants.

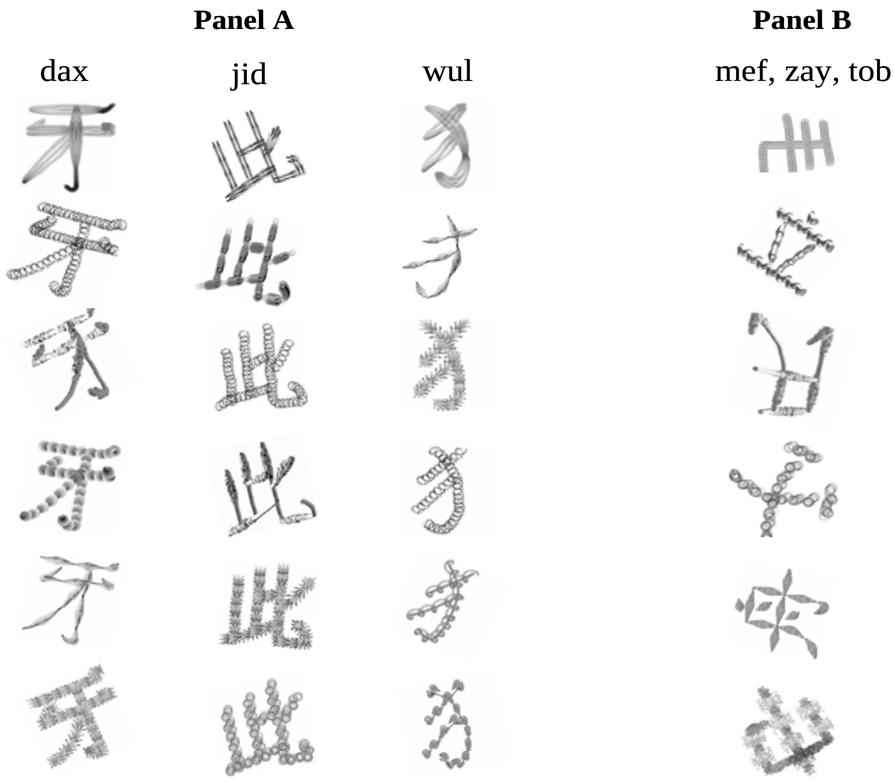


Figure 12. Examples of category members for the three acquired categories, under the nonsense syllable that served as each category's name (Panel A). Examples of non-category stimuli under the nonsense syllables that were paired randomly with these stimuli (Panel B).

Participants learned the three categories in practice sessions outside the scanner. In a first session, they learned the categories using a variety of standard learning methods. During paired associate presentation, participants studied the name of a category shown together with a picture of a category member. During paired associate matching, participants were shown a name with three pictures and had to pick which picture was a member of the named category. During drawing, participants were shown a category name and had to draw a category instance. During paired associate verification, participants were shown a name followed by a picture and had to say whether the picture was an instance of the named category. During paired associate naming, the order was reversed, and participants received a picture followed by a name, and had to say whether the name correctly named the preceding picture.

As described in the next section, participants also viewed what we will call "control" stimuli, which came from no category. Panel B of Figure 12 presents examples of control stimuli. As can be seen, these stimuli did not share a common form, such that subsets of them did not form categories. The form in a given control stimulus was never repeated in another control stimulus. Additionally, control stimuli were associated with three nonsense syllables (also shown in Panel B). On a given control trial, one of these nonsense syllables was randomly associated with a control stimulus. Across control trials, the nonsense syllables associated with them became familiar, but because the control stimuli did not have a categorical structure, these three nonsense syllables never became associated with a concept for a particular kind of visual structure.

No category member was ever repeated during the entire experiment, across all training phases and the scanning phase. For a familiar category, each of its members was only shown once. Each control stimulus was also shown just once. The only stimuli that repeated across experiment were the six nonsense syllables, three for the familiar categories and three for the control stimuli.

Once participants had learned the categories, they practiced the specific tasks that they would have to perform in the scanner (the next section describes these tasks in detail). Specifically, participants practiced exactly the same kinds of runs that they would perform the next day in the scanner. After the first practice session, participants took a study sheet home that summarized the categories, and that provided them with additional drawing exercises to perform on their own. The next day, just before being scanned, participants performed additional practice runs outside the scanner. They then performed the critical runs in the scanner.

2. Types of trials. This experiment used 10 different types of trials to address the hypotheses of interest. Table 4 lists these 10 trial types and their characteristics. As Table 4 illustrates, the types of trials fall into two general groups: verification trials and naming trials. Each group is addressed in turn.

Trial Type	**Trial Segments**	**Frequency**
Familiar verification trial	name – picture – response	36
Familiar verification catch trial	name – picture	12
Unfamiliar verification trial	name – picture – response	36
Unfamiliar verification catch trial	name – picture	12
Verification catch trial	name	24
Familiar naming trial	picture – name – response	36
Familiar naming catch trial	picture – name	12
Unfamiliar naming trial	picture – name – response	36
Unfamiliar naming catch trial	picture – name	12
Naming catch trial	picture	24

Table 4. Types of trials, their trial segments, and their total frequency in the experiment.

On verification trials, a name was presented, followed by a picture (see Table 4). A participant's task was to verify whether the name and picture matched (i.e., by making a binary response of match vs. mismatch). Thus, verification trials contained three components that occurred in sequence: name, picture, response. Verification trials could be familiar or unfamiliar. On familiar verification trials, the name was the name of a familiar category, and the picture was also from a familiar category. Half the time, the name and picture matched (i.e., they were from the same category), and half the time they did not (i.e., they were from different categories). On unfamiliar verification trials, a nonsense syllable not associated with a category was followed by a picture that did not belong to one of the three categories (i.e., a control stimulus). On these trials, participants pressed a third response key to indicate that the name and picture were not from a familiar category.

On naming trials, a picture was presented, followed by a name (see Table 4). A participant's task was to assess whether the picture was named correctly (i.e., by making a binary response of match vs. mismatch). Thus, naming trials contained three components that occurred in sequence: picture, name, response. Naming trials could be familiar or unfamiliar. On familiar naming trials, the picture was from a familiar category, and the name was the name of a familiar category. Half the time, the picture and name matched (i.e., they were from the same category), and half the time they did not (i.e., they were from different categories). On unfamiliar naming trials, a picture that did not belong to one of the three categories was followed by a nonsense syllable that was not associated with a category. On these trials, participants pressed a third response key to indicate that the picture and name were not familiar.

As Table 4 further illustrates, additional verification and naming trials served as *catch trials*. The catch trials allowed us to deconvolve the BOLD responses for the three individual components of the verification trials (name, picture, response). The catch trials similarly allowed us to deconvolve the BOLD responses for the three individual components of the naming trials (picture, name, response). In an event related design, such as this one, BOLD responses can usually be deconvolved for different events only if they are relatively far apart in time, with the temporal intervals between them varying randomly. This makes it impossible to present two stimuli in a fixed sequence, with a short temporal interval between them. For example, it would normally not be possible to isolate individual events in the verification and naming trials described above (names, pictures, responses), because they require short fixed sequences. However, researchers have previously figured out how to deconvolve *two* adjacent events, separated by a short fixed time interval (Ollinger, Shulman, & Corbetta, 2001a,b).

Our contribution to this methodology is that we have figured out how to deconvolve *three* sequential events separated by short fixed time intervals. To our knowledge, no one has ever described or performed a design that accomplishes this. The design in Table 4, however, does. It allows us to extract the BOLD response for the first event in a sequence (e.g., the name in a verification trial), the second event in the sequence (e.g., the picture in a verification trial), and the third event in the sequence (e.g., the response in a verification trial). The three kinds of catch trials for the verification trials create contrasts with the full verification trials that make it possible to extract the BOLD response for each of the three components. Analogously, the three kinds of catch trials for the naming trials create contrasts with the full naming trials that make it possible to extract the BOLD response for each of their three components. In general, this kind of design could be of considerable use as researchers attempt to implement complex cognitive tasks in neuroimaging research.

Finally, the timing of the trials proceeded as follows. A fixation cross appeared between trials for a random interval that ranged from 2.5 to 25 sec, thereby implementing the optimal conditions for an event related design. When a trial occurred, either a word or picture appeared for 2.5 sec in the center of the screen. If this was a name-only or picture-only catch trial, the fixation cross reappeared, and the participant made no response. If another event occurred, it was again always a picture or name presented for 2.5 seconds, again in the center of the screen. If this was a name-picture or picture-name catch trial, the fixation cross reappeared, and the participant made no response. If this was not a catch trial, then a down arrow appeared, indicating that the participant was to make a match, mismatch, or unfamiliar response on the button box as quickly as possible while maintaining accuracy. Participants were not allowed to respond until they saw the down arrow. After the participant made a response, the fixation cross reappeared. Following another variable fixation interval, another trial began. Participants knew that catch trials would occur and practiced them in the runs performed outside the scanner.

3. Results. Brain activations were computed for 14 participants who exhibited low amounts of movement in the scanner and high behavioral performance. We used a threshold of $p < .001$ for individual voxel significance. We also applied a cluster size threshold that varied by condition as function of smoothness to produce an overall corrected significance level of $p < .05$. Clusters significant by these criteria ranged in size from approximately 12 to 35 contiguous functional voxels (1.7 x 1.7 x 3.0 mm). Random effects ANOVA were performed on contrasts that tested hypotheses of interest.

Figure 13 illustrates the logic behind the contrast of primary interest. As described at the top, a category name on verification trials (e.g., dax) should activate its meaning. Because PSS predicts that this meaning will be represented as a simulation in visual areas, we predicted that category names would activate visual areas, in particular, the ventral stream (and possibly the dorsal stream). Once such a simulation exists, participants can then compare it to visual perception of the subsequent exemplar to see if they match, which is what is required to make a correct response.

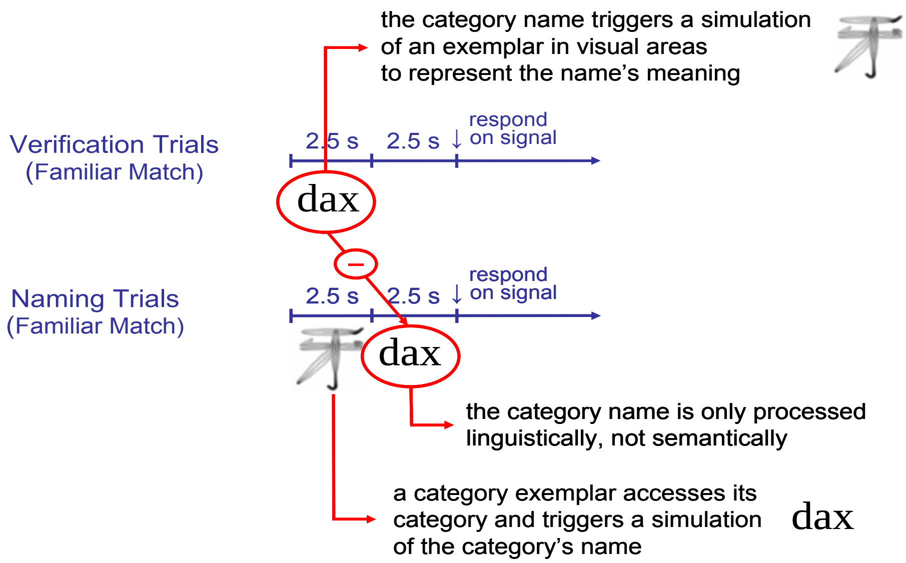


Figure 13. Process model underlying the contrast of primary interest in the fMRI category learning experiment.

As illustrated at the bottom, a category name on naming trials should only produce superficial linguistic processing, not activation of its meaning. The logic behind this is as follows. When an exemplar is presented first, it activates the name of its category. The participant then waits until a name is actually presented and assesses whether it is the predicted name. To make this assessment, it is only necessary to process the name as a linguistic object that can be compared to the anticipated category name. It is not necessary to activate its meaning.

Most importantly, by subtracting brain activations for the names presented second on naming trials from brain activations for the same names presented first on verification trials, areas of the brain that represent meanings of the names can be isolated. Of interest is whether these areas reside in the ventral and dorsal streams. As predicted by PSS, participants should simulate exemplars in the visual system to represent the meanings of the category names

Figure 14 shows results from this contrast. As can be seen, huge activations reside in both the ventral and dorsal streams. Although exactly the same stimuli occurred in the two conditions contrasted—the familiar category names—large differences in activation occurred. Because the meanings of the category names should be active when the names are presented on verification trials but not on naming trials, the activations in Figure 14 represent where these meanings are stored in the brain. Consistent with PSS, the meanings of category names are simulated in the relevant modality-specific systems.

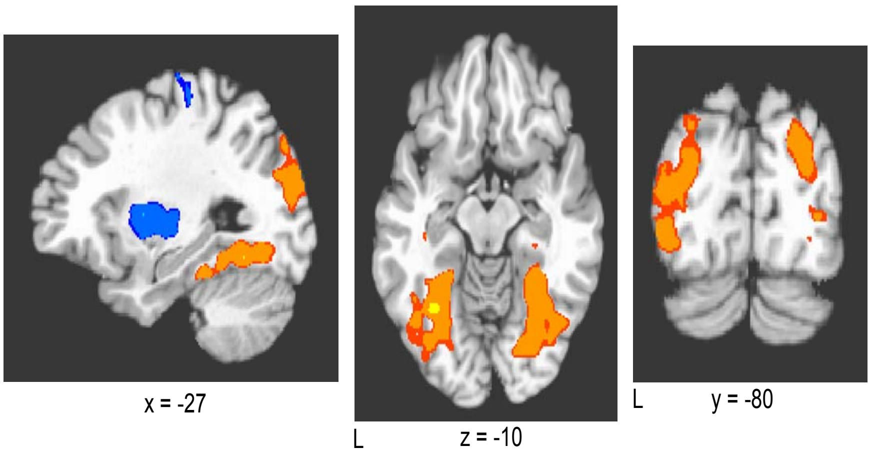


Figure 14. Activations from the fMRI experiment on category learning produced by subtracting activations for category names on naming trials from activations for category names on verification trials

One other contrast from this experiment—illustrated in Figure 15—is also of interest. In this contrast, activations for exemplars presented on first on naming trials are subtracted from activations for names presented first on verification trials. Exemplars presented on first on naming trials should produce activations associated with visual processing of the exemplars and with categorizing them. In contrast, names presented first on verification trials should produce activations associated with recognizing the names and activating their meanings. PSS predicts that there should be considerable overlap between these two sets of activations, given that the names' meanings should be simulations of the visual and categorical processing associated with the exemplars.

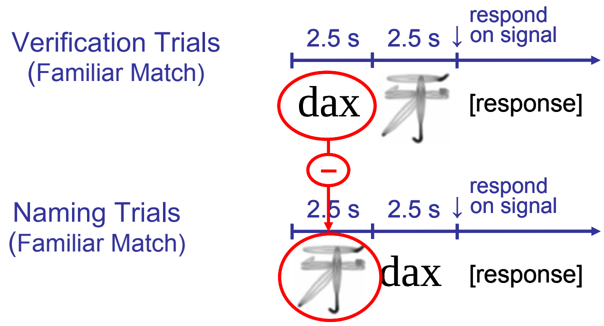


Figure 15. An additional contrast of interest in the fMRI category learning experiment.

As Figure 16 illustrates, this hypothesis was supported. All of the blue activations in this figure are areas that were *more* active for the exemplars than for the names. Of considerable interest is the high overlap of these activations with those in Figure 15, which again represent the meanings of the names. The high overlap between these activations indicates that the name meanings activated the same areas as the exemplars, but not as highly—these overlapping activations were higher for the exemplars, as the contrast in Figure 16 indicates. This pattern is highly consistent with many findings in the mental imagery literature, where mental imagery of a stimulus activates the same areas associated with perceiving the stimulus, but to a lesser extent. As Figures 15 and 16 show together, the names here activated the same areas as the exemplars, but to a lesser extent. This pattern further supports the conclusion that simulations of exemplars constitute the meanings of category names.

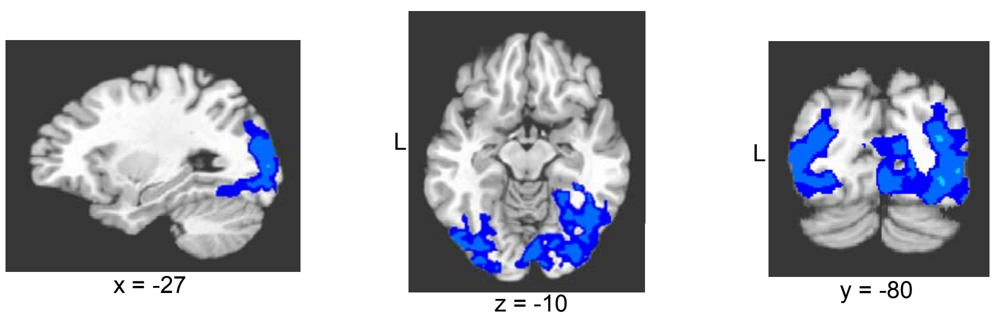


Figure 16. Activations from the fMRI experiment on category learning produced by subtracting activations for exemplars on naming trials from activations for category names on verification trials. Blue areas are brain regions that were more active for exemplars than names. Note the high overlap between these activations for exemplars and the activations for name meaning in Figure 15.

The differences—not just the similaritities—between the activations in Figures 15 and 16 are also of interest. As can be seen, posterior activations in the early visual system only occurred for the exemplars in Figure 16 but not for the names in Figure 15, indicating greater visual processing associated with the exemplars. Conversely, anterior activations in association areas of the temporal lobes occurred for the names in Figure 15 but for the exemplars in Figure 16, indicating greater linguistic processing associated with the names. Nevertheless, these differences are much smaller than the overlapping areas of activation, again supporting the conclusion that simulation underlies meaning.

4. Conclusions and further studies. The paradigm developed here offers a new way to identify the representations of categories, and to assess whether these representations conform to the predictions of PSS. Unlike previous paradigms in grounded cognition, this one allows for careful control over novel categories under controlled learning conditions, rather than relying on pre-existing natural categories. Because of this greater control, we should be able map out the linguistic and visual circuits that represent simple visual categories, such as those that simple computational agents encounter. Finally, this paradigm offers a new double deconvolution procedure that makes it possible to isolate components in fixed three-event trial sequences. Further experiments using this paradigm are planned.

III. Evidence for Simulation in Conceptual Combination

Two lines of research were developed under this DARPA contract to assess whether the symbolic operation of conceptual combination is grounded in simulation. One line of research uses a behavioral paradigm, and the other uses fMRI. Both paradigms are novel, not having been used by other researchers or ourselves prior to DARPA funding. Both paradigms offer much potential for studying the fundamental process of conceptual combination and for assessing theoretical accounts of it. Each paradigm, along with the results obtained from it to date, is addressed in turn.

A. Behavioral Evidence for Simulation in Conceptual Combination

This next line of work uses a behavioral paradigm to assess whether people combine the meanings of words in a noun phrase by simulating their individual meanings and then combining these simulations into a larger simulation. If people do form conceptual combinations this way, it would have implications for building artificial agents who must combine the meanings of words in simple commands (e.g., “push the button”, “lift the block”). Rather than comprehending such expressions by combining amodal symbols for the individual words, artificial agents could comprehend these expressions instead by combining simulations of word meanings.

The experiments in this line of research build on existing paradigms that assess whether the meanings of individual words are grounded in simulation. Essentially, we are extending these paradigms so that they can be used to assess whether the meanings of *multiple-word* expressions are also grounded in simulation. Thus, the innovation in these experiments is adapting existing paradigms for the study of individual words so that they can be used to study the conceptual combination of multiple words.

In this next line of research, we assessed conceptual combination in simple noun phrases, such as *sky diver*. All of the simple nouns phrases used contain a modifier (*sky*) and a head noun (*diver*). Our hypothesis is that participants simulate the meanings of the modifier and the head noun to combine them conceptually. If this hypothesis is correct, then perceptual variables such as height should affect this process.

1. Method. On each trial, participants viewed a fixation cross on the computer screen and pressed a foot pedal when ready to initiate the trial. Following a 1 sec blank screen, a modifier appeared for 1 sec, followed by a 500 ms blank screen, and then a head noun. On seeing the head noun, the participant pressed a button as quickly as possible to indicate whether the noun phrase referred to something that is real (e.g., *sky diver*) or unreal (e.g., *saxophone bee*).

One of the key manipulations in the experiment concerned the vertical height of the head noun on the screen. On half the trials, the head noun appeared at the top of the screen; on half the trials, the head noun appeared at the bottom. For example, participants received the word *sky* centered in the screen, followed by the word *diver* either at the top or the bottom of the screen. A given participant only ever saw a particular head noun at the top or bottom, not both. Across trials, however, a participant saw many head nouns at both positions. The position of a head noun was counter-balanced between participants, such that the same head noun occurred equally often in both positions.

The spatial arrangement of the modifier and head noun leads to the following prediction. According to the PSS account of conceptual combination, people immediately begin simulating the modifier’s meaning (e.g., *sky*) as soon as they read its word (cf. Marslen-Wilson & Tyler, 1980). Thus a simulation should have begun to develop before the head noun appears at the top or bottom of the screen (e.g., *diver*). Once participants have comprehended the head noun, they must shift attention to the relevant part of the modifier simulation to combine the head noun simulation with it. To combine a *diver* simulation with the *sky* simulation, for example, participants must internally shift attention up in the simulated *sky*, given that sky divers typically occur at a high spatial position.

Most importantly, PSS predicts that the location of the head noun on the screen should interact with the internal shift of attention needed to combine simulations. When the word "diver" appears at the top of the screen, participants must shift their gaze upward to process it. Because this external visual shift is consistent with the internal shift that combines the modifier and head noun simulations, the external shift positions attention in the optimal location for combining simulations. Conversely, when the word "diver" appears at the bottom of the screen, attention must shift down to read it, thereby drawing visual attention away from the internal position required for combining simulations. A subsequent shift to the top of the *sky* simulation must follow, thereby slowing response time.

Amodal theories do not naturally predict this effect. According to these theories, participants retrieve amodal symbols for *sky* and *diver,* and then combine them in an amodal relational structure, such as *IN (diver, sky*). The screen position of the second word should not affect the process of combining these amodal symbols. Nothing in the syntactic operation of combining two symbols has anything to do with the height of an internal symbol or the height of a word in the display (imagine a computer combining symbols obtained from the top or bottom of the screen). Conversely, PSS naturally predicts and explain effects of this manipulation. Because the brain is running a visual simulation that has a vertical dimension, operations on the simulation are affected by the vertical position of selective attention. If the head noun in the display draws attention to the wrong vertical position, this should interfere with the role of selective attention in combining simulations.

To control for vertical bias associated with particular head nouns (e.g., *diver*), each head noun was also combined with a modifier that predicts faster processing when the head noun appears at the bottom of the screen. On other trials, for example, *scuba* was the modifier presented before *diver*, where *diver* again appeared at the screen's top or bottom. If participants run simulations to combine to represent *scuba diver,* then a consistent shift of attention down when "diver" appears a the bottom of the screen should produce faster responses than when "diver" appears at the top.

Each of 48 critical head nouns occurred in four types of trials, as illustrated by *sky diver* (screen top), *sky diver* (screen bottom), *scuba diver* (screen top), *scuba diver* (screen bottom). Table 5 illustrates further examples. A given participant received 12 trials of each type, but only received a given modifier and head noun once. Materials were counter-balanced across participants so that each type of trial for a head noun occurred equally often. Head nouns were selected that varied in vertical height, with some typically occurring in high, intermediate, or low positions relative to the perceiver (e.g., *head, cushion, frog*).

In addition to the 48 critical trials, a participant received 240 filler trials to mask the critical materials and the purpose of the experiment. These fillers included 96 other "real" trials that had nothing to do with height, so that other relations between modifiers and head nouns were salient. On the remaining 144 trials, participants received modifiers and head nouns that did not refer to anything real. On 48 of these trials, height was a relevant relation (e.g., *sun jeep, deep copy*), such that a height relation could not be used as a cue for responding "real." Table 5 presents examples of filler trials.

Critical Real Items		
High Focus	**Low Focus**	
giraffe head	lizard head	
monster truck	toy truck	
basketball net	tennis net	
watertower tank	gasoline tank	
should pad	sandal pad	
climbing squirrel	digging squirrel	
	Fillers	
Non-Height Real	**Height Non-Real**	**Non-Height Non-Real**
sugar crystal	comet harpsichord	ear aluminum
ranch brand	helicopter hedge	butterfly ambulance
calculator cover	skyscraper nutmeg	fridge bracelet
island beach	tunnel eclipse	pear canal
apricot farm	valley helmet	minnow car
clock gear	root jacket	tweezers cauliflower

Table 5. Examples of the materials from the behavioral experiment on conceptual combination.

2. Results. Based on the 12 participants run in this experiment so far, vertical position appears to have an effect on conceptual combination (Figure 17). Participants were 114 ms faster when the vertical position of the head noun was consistent with the vertical position of the head noun's meaning than when the two heights were inconsistent.

Consistent with PSS, participants appeared to perform conceptual combination by combining simulations for the modifiers and head nouns. It is difficult to reconcile these results with theories that assume manipulation of amodal symbols underlies conceptual combination (as when computers perform this process). Instead, humans appear to ground conceptual combination in modality-specific simulations.

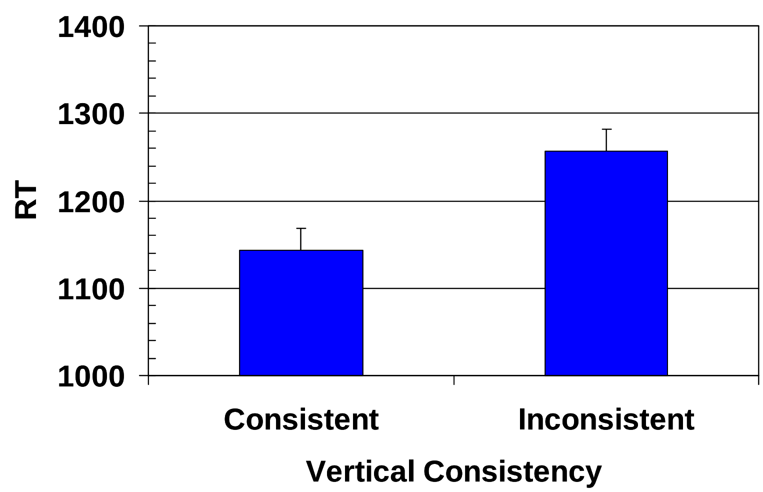


Figure 17. Average RT to determine whether a noun phrase refers something that is real or not real as a function of whether the screen position and meaning of the head noun were consistent or inconsistent in vertical position.

3. Failures to replicate. Based on the success of this first experiment, Aron Barbey performed five further experiments for his dissertation to further asssess the role of simulation in conceptual combination. Barbey's work on these experiments has been completed, and his dissertation was defended in July 2007. To foreshadow, none of these experiments confirmed the simulation hypothesis, unlike the experiment just reported. This is the only project of the six projects reported here that did not consistently produce positive results. We are still in the process of trying to understand what happened with these experiments.

The first three experiments of Barbey's dissertation used the same materials as the experiment just described, where the height of a noun phrase was manipulated by using either a high or low modifier for the same head noun (see Table 5). All that differed was the task that participants performed on these materials. Thus, these experiments replicate the materials of the experiment just reported but with different tasks.

In the first experiment of Barbey's dissertation, participants read sentences one word at a time in the center of the screen, pressing a response button after reading each word, so that they could procede to reading the next word. To ensure that participants processed the words in each sentence deeply, participants had to indicate whether the sentence made sense after reading its final word. For example, participants read, "A hanging rug could have woven fibers" and then indicated (via a button press) that the sentence made sense, as opposed to "A flying duck could have feathered talons," which did not.

The first noun phrase in each sentence was of primary interest. These first noun phrases where exactly the same noun phrases as used in the earlier experiment. To manipulate display height, the head noun in these critical noun phrases appeared either at the top or bottom of the screen, rather than in the center (most other words appeared in the center). As in the previous experiment, the display height of a head noun was either consistent or inconsistent with the meaning of the noun phrase. For the noun phrase, "flying duck," duck could appear at the top of the screen, consistent with the meaning of the noun phrase, or at the bottom of the screen, inconsistent with the meaning. Furthermore, the same head noun was associated with another modifier that changed the height of the noun phrase's meaning. For example, "duck" was also paired with "swimming" to create "swimming duck." Again consistency was manipulated by presenting "duck" in this phrase at the top (inconsistent) or bottom (consistent) of the screen. As described earlier, a given participant never saw both noun phrases that contained the same noun, although height and consistency were fully counter-balanced between participants across versions.

Again, many filler sentences were used. Rather than the head noun of the first noun phrase varying in height, however, a later word in these filler sentence varied, appearing either at the top or bottom of the screen. These later words in the filler sentences that varied in height masked the height variation of interest, namely, the height variation of the first noun phrases that contained the critical materials.

Panel A of Figure 18 shows the results of this experiment. As can be seen, there was no consistency effect. Time to read the a critical head noun was unaffected by whether its position at the top or bottom of the screen was consistent with the meaning of the noun phrase in which it appeared.

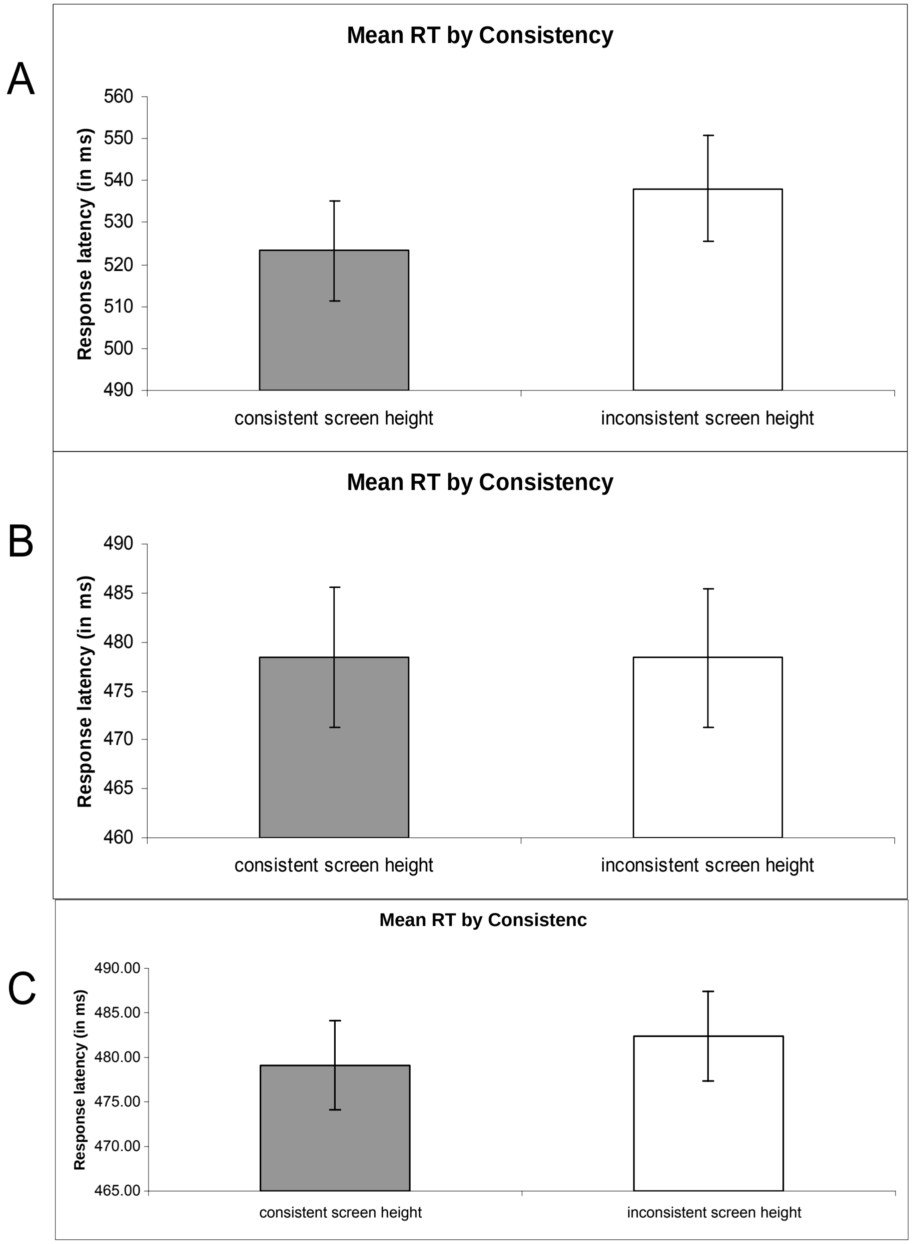


Figure 18. Results from Experiment 1 (Panel A), Experiment 2 (Panel B), and Experiment 3 (Panel C) of Aron Barbey's dissertation that assessed height effects in conceptual combination.

The second experiment of Barbey's dissertation was the same as the first, except for a slight difference in task. Again, participants read words one at a time in the center of the screen, except that now they appeared at a fixed rate (600 ms per word). Furthermore, all words appeared in the center of the screen, with none varying in height. Finally, the major change was that a dot appeared 200 ms after one word in the sentence at either the top or bottom of the screen. On detecting the dot, participants had to press a key to indicate its presence. Again, participants indicated whether the sentence made sense after reading it, thereby ensuring deep processing.

Of primary interest was whether the height of the dot interacted with the meaning of the critical noun phrases. For all of these noun phrases, the dot always appeared 200 ms after the head noun had been presented (for the filler sentences, the dot always appeared after later words in the sentence to mask the critical materials). Most importantly, the height of the dot was either consistent or inconsistent with the meaning of the noun phrases. After reading “flying duck,” a dot appearing at the top of the screen was consistent, whereas a dot appearing at the bottom of the screen was inconsistent. Conversely, after reading “swimming duck,” a dot appearing at the top of the screen was inconsistent, whereas a dot appearing at the bottom of the screen was consistent. If participants simulate the meanings of “flying duck” and “swimming duck” to represent them, then the height of the dot should interact with processing the meaning of the noun phrase.

As Panel B of Figure 18 shows, however, there was no such effect. Time to detect a critical dot was unaffected by whether its position at the top or bottom of the screen was consistent with the meaning of the noun phrase after which it appeared.

The third experiment of Barbey’s dissertation was the same as the second except that the dot was place by either an X or O at the top or bottom of the screen, and participants had to categorize it as an X (left button) or O (right button) instead of simply detecting the stimulus, as was the case for the dot in the previous experiment. Of interest here was whether a deeper categorization task (X vs. O) would produce a height consistency effect, relative to the simpler perceptual task of simply detecting a dot.

As Panel C of Figure 18 shows, however, there was again no effect. Times to categorize Xs and Os were unaffected by whether their position at the top or bottom of the screen was consistent with the meaning of the noun phrase after which they appeared.

It occurred to us that our intuitive classification of height in originally sampling the modifiers, head nouns, and noun phrases might have been inaccurate. To assess this possibility, Experiment 4 of Barbey’s dissertation carefully scaled these materials for height. These scalings indicated, however, that our original sampling was highly accurate. Modifiers, nouns, and noun phrases that we had assigned to high and low conditions were indeed high or low, respectively, as judged by this independent group of participants. Furthermore, we used these scalings in further regression analyses to more rigorously assess the ability of height to predict performance in the previous three experiments. Again, we found no consistent effects of height consistency, for the modifiers, nouns, or noun phrases.

The final experiment of Barbey’s dissertation, Experiment 5, took another approach to examining the role of consistency in conceptual combination. In an initial learning phase, participants associated individual words with pictures of their referents. Figure 19 presents examples of these word-picture pairings. When a word was associated with two pictures (e.g., “cake”), participants only studied one of two pictures, not both (see Figure 19). The purpose of this manipulation is explained later. Participants studied each word-picture pair a total four times. Thus, the words and pictures were well associated by the end of the initial study phase.

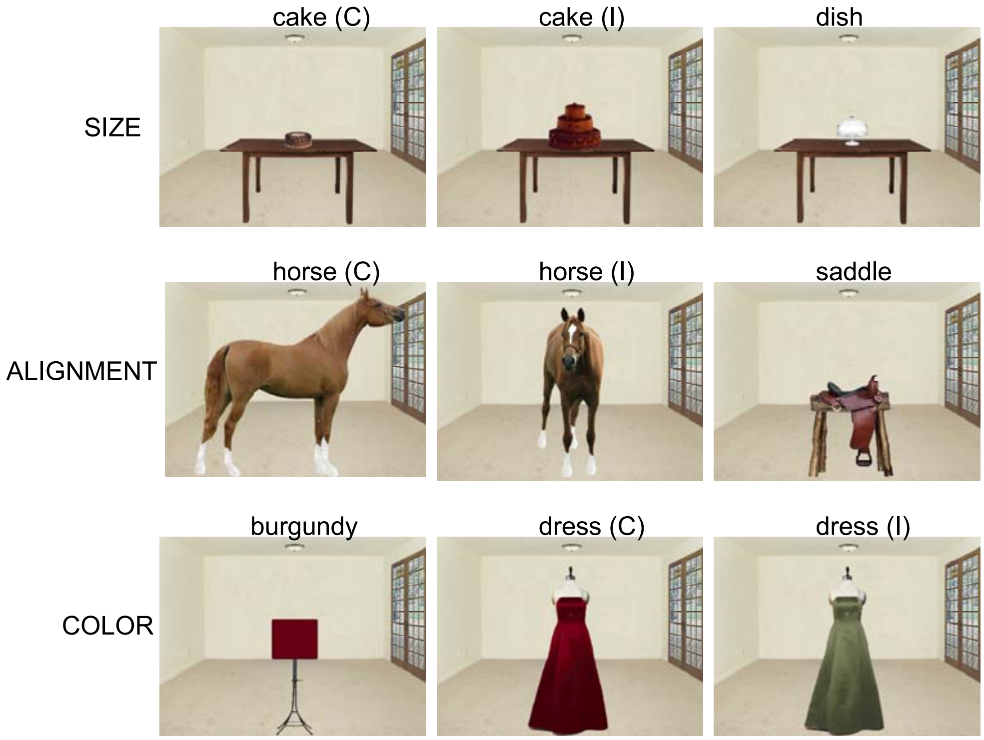


Figure 19. Examples of the materials from Experiment 5 of Aron Barbey's dissertation. C is a consistent modifier in the size and alignment conditions, or a consistent noun in the color condition. I is an inconsistent modifier in the size and alignment conditions, or an inconsistent noun in the color condition.

Following the study phase, participants received noun phrases, each consisting of two words whose pictures had been studied earlier (e.g., "cake dish" made up from "cake" and "dish"). Participants comprehended each noun phrase and then judged the pleasantness of its meaning. Of primary interest was the time to comprehend the noun phrase before entering a judgment. Similar to Experiments 1, 2, and 3, Experiment 4 asesssed the effect of consistency on conceptual combination. Unlike these previous experiments, however, height was not the critical factor. Instead, the effects of three new factors on consistency were assessed: size, alignment, and color.

First consider size. As Figure 19 illustrates, the cake that a given participant studied initially could have been small or large. As can be seen, the small cake fits in the dish on the right of Figure 19, but the large cake does not. If perceptual simulation underlies conceptual combination, then when participants activate the meanings of "cake" and "dish" to compute the meaning of "cake dish," they should activate simulations of the cake and dish pictures studied earlier, and then attempt to integrate these simulations with the cake inside the dish. Participants who studied the small cake earlier should make these judgments faster than subjects who studied the large cake, given that the small cake's size is consistent with the size of the dish, thereby making it easy to integrated simulations of them. Conversely, participants who studied the large cake earlier, should have more difficulty combining simulations because the large cake does not fit inside the dish.

As Figure 19 illustrates, consistency was manipulated similarly for alignment and color. For alignment, one of the two pictures associated with modifier could be easily aligned with the picture shown for the head noun, whereas the other picture associated with the modifier could not (e.g., a horse either aligned or not aligned with a saddle). Thus, when participants later received "horse saddle," they should be faster when the pictures were aligned earlier than when they were not. For color, a picture of a color patch (e.g., burgundy) could be the same as the color of the pictured object shown for the head noun or different (e..g, a burgandy dress or an olive dress). Thus, when participants received "burgundy dress," they should be faster when they had seen a burgundy dress earlier than when they hadn't.

Figure 20 shows the results of Experiment 5. As can be seen, consistency effects again failed to occur. The time to comprehend a noun phrase prior to judging its pleasantness was unaffected by the consistency of the pictures for the modifier and head noun studied earlier. We also examined several other measures of noun phrase processing, including average pleasantness and later memory of the noun phrases. With a few exceptions, these measures also did not show consistency effects.

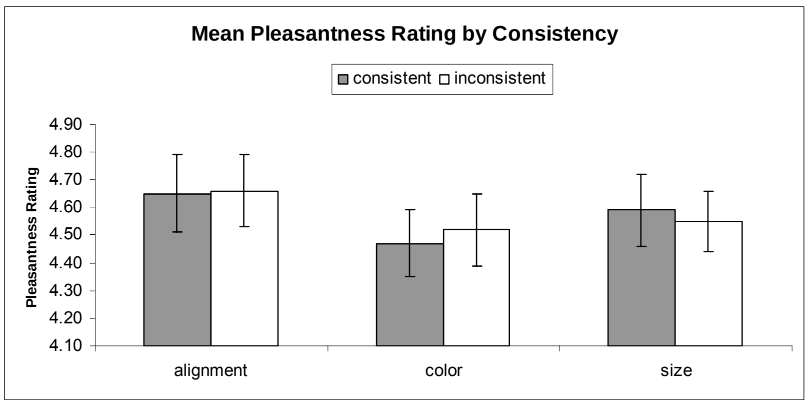


Figure 20. Results from Experiment 5 of Aron Barbey's dissertation that assessed consistency effects in conceptual combination for alignment, color, and size.

4. Issues and further research. We are quite puzzled about our failure to find consistent simulation effects for most of the experiments in this project. As will be seen in the next project, we found large simulation effects in an fMRI experiment. Furthermore, other behavioral research on conceptual combination has found simulations effects (Estes, Verges, & Barsalou, submitted; Wu & Barsalou, in preparation). Furthermore, much other research has found height effects in the processing of individual words (e.g., Meier & Robinson, 2004; Shubert, 2005). Thus, there appears to already be a significant amount of evidence that simulation underlies conceptual combination. In addition, consistency effects of this sort are widespread in the literature (for a review, see Zwaan & Madden, 2005).

A variety of factors could have prevented simulation and consistency effects here. There could be subtle details of our procedures that mitigated effects (e.g., instructions, tasks, fillers). Given that many of these studies were run toward the end of the semester, we could have had unusually noisy samples of participants. Another possibility is that egocentric vs. allocentric spatial processing was not controlled in these studies and could have been a factor. Although most studies like these do not control for this type of processing, increasing work suggests its importance in perception and action. Given that we assume cognition utilizes perceptual and motor processes, egocentric vs. allocentric strategies could be an important issue for future research to explore.

More generally, much further work remains to be done that assesses the specific processing mechanisms underlying conceptual combination, not only in simple noun phrases, but in all the complex constructions that underlie text meaning. In our opinion, understanding these processes is one of the most important issues facing cognitive science and cognitive neuroscience.

B. fMRI Evidence for Simulation in Conceptual Combination

This next line of work assesses whether neural evidence corroborates the behavioral finding that conceptual combination is grounded in modality-specific simulation. If it is, then activation related to conceptual combination should occur in the brain's modality-specific systems as people perform conceptual combination.

Many researchers believe that relations are central to the process of combining symbols, as when people combine the meanings of a modifier and head noun in a noun phrase (e.g., Gagne & Shoben, 2002; Levi, 1978). In *floor heater*, for example, a *location* relation specifies that the heater is *on* the floor. Linguists and psychologists have identified a variety of important relations that frequently structure conceptual combinations.

Nearly all existing theories assume that amodal structures represent these relations, as for *ON* (*x*, *y*), where *x* and *y* bind to the head noun and modifier, respectively, to form *ON (x=heater, y=floor)*. Thus these theories predict that the brain's modality-specific systems should not be central to representing relations in conceptual combinations. Instead, amodal structures somewhere else in the brain represent them. Alternatively, PSS proposes that these relations are grounded in the modalities, not outside them.

1. Method. Participants received four types of trials in an fMRI scanner, after practicing these trials outside the scanner beforehand. Table 6 illustrates the four trial types, which occurred in an event-related design. On most trials, participants received a modifier in the center of the screen for 1 sec, followed by a blank screen for 3 sec. A head noun then appeared for 1 sec, also in the center of the screen, again followed by a blank screen for 3 sec. Thus, the basic trial format was for a noun phrase to be presented one word at a time at a 4 sec SOA.

Trial Type	Motion Modifiers	Location Modifiers	Mental State Modifiers
Conceptual Combination Trial	swimming– / father swaying– / oak soaring– / balloon	auditorium– / piano ocean– / shrimp closet– / gnat	distressed– / reverend gloomy– /dog pleasing– / cloves
Independent Words Trial	swimming. / father swaying. / oak soaring. / balloon	auditorium. / piano ocean. /shrimp closet. / gnat	distressed. / reverend gloomy. / dog pleasing. / cloves
Combination Catch Trial	rolling– falling– vibrating–	apartment– mountain– attic–	persuasive– merry– delightful–
Independent Catch Trial	rolling. falling. vibrating.	apartment. mountain. attic.	persuasive. merry. delightful.

Table 6. Examples of trials from the fMRI experiment on conceptual combination. Modifiers referred to motions, locations, or mental states. On conceptual combination trials, the modifier appeared in the middle of the screen. The – after the modifier indicated that participants were to wait until the noun was presented before evaluating the familiarity of the entire noun phrase. The / indicates that the modifier disappeared and then the subsequent head noun appeared. On independent word trials, participants first judged the familiarity of the modifier, and then judged the familiarity of the head noun. The . indicated that participants should evaluate each word separately, rather than only evaluating the entire noun phrase. On catch trials, participants prepared to evaluate either the noun phrase or the head noun after the head noun appeared, but it never did.

On catch trials, modifier appeared, but then a head noun never followed. As described earlier, these catch trials allowed us to deconvolve the activations for the modifiers and head nouns, even though the modifiers and head nouns occurred close in time together, always a fixed interval apart. In this experiment (as opposed to the previous one), we only performed a single deconvolution, which is a standard operation in the literature.

To separate brain activations for conceptual combination from brain activations for individual words, we manipulated the following variable. On some trials, participants evaluated the entire noun phrase as a unit. On other trials, participants evaluated each word in the noun phrase separately. Thus, participants perceived exactly the same word sequence in both conditions (a modifier and then a head noun), but processed them differently, either as a noun phrase or as individual words.

Table 6 illustrates how this manipulation was implemented. When the modifier appeared, it was followed by either a dash (–) or a period (.). When a dash followed the modifier, this indicated that the participant was to judge the modifier together with the noun as a noun phrase. In other words, the participant had to wait until the head noun appeared before making a judgment. The participant's task was to judge whether the noun phrase was very common, somewhat common, or rare, by pressing one of three buttons on a response box.

Conversely, when a period followed the modifier, this indicated that the participant was to judge the modifier's familiarity first, and then judge the head noun's familiarity separately, after the head noun appeared later. Thus, the participant judged the familiarity of each word in sequence, rather than the familiarity of the noun phrase as a whole. This manipulation allowed us to assess the brain areas unique for performing conceptual combination above and beyond the areas required for processing individual words. While the stimulus presentation was identical in the two conditions (except for the dash vs. period), the processing required varied.

Table 6 also illustrates that participants received two kinds of catch trials, where a head noun never followed the modifier. On some catch trials, participants believed that they were supposed to evaluate the entire noun phrase after the head noun appeared (although it never did). On other catch trials, participants evaluated the modifier, and then waited to evaluate a head noun that did not appear.

To assess whether conceptual combination is grounded in modality-specific simulation, we manipulated whether the modifier referred to a motion, location, or mental state. Table 6 provides examples of modifiers from all three domains.

Participants received a total of 180 trials on which both a modifier and head noun were present, and they recreived a total of 60 catch trials. Half of the trials in each group were processed as conceptual combinations, and half were processed as independent words. Thus, 90 of the complete trials were processesd as conceptual combinations, as were 30 of the catch trials. Orthogonally, a third of the modifiers came from each of the three domains (motions, locations, mental states). Thus, 60 of the modifiers on the complete trials were from each domain, as were 20 of the catch trials.

No modifier or head noun ever repeated. The head nouns were carefully controlled so that their semantics, category membership, and typicality were the same for each of the three modifier domains. This control of the head nouns is critically important for the contrasts performed in the later analysis. First, the 60 head nouns used to construct the noun phrases in each modifier domain were equivalent in Kucera-Frances word frequency. Second, the head nouns were also equivalent in terms of their category membership. Half the nouns in each domain were animate, and half were inanimate. Within the animate and inanimate groups for each domain, head nouns were drawn from the same semantic categories, and they had equivalent typicality levels within these categories. For example, the animate nouns in each domain were drawn equivalently from animate categories such as fruit, vegetables, and mammals, and had equivalent typicality levels. Similarly, inanimate nouns were drawn equivalently from inanimate categories such as furniture, clothing, and weapons, and had equivalent typicality levels. As a result of this careful and precise sampling process, the head nouns used to construct nouns phrases in the three modifier domains were as equivalent in terms of frequency and semantics.

2. Results. Brain activations were computed for 15 participants who exhibited low amounts of movement in the scanner and high behavioral performance. We used a threshold of $p < .001$ for individual voxel significance. We also applied a cluster size threshold that varied by condition as function of smoothness to produce an overall corrected significance level of $p < .05$. Clusters significant by these criteria ranged in size from approximately 12 to 30 contiguous functional voxels (3 x 3 x 3 mm). Random effects ANOVA were performed on contrasts that tested hypotheses of interest.

Figure 21 shows strong modality-specific effects when participants judged the modifiers independently (i.e., participants judged the familiarity of the modifier and head noun separately). Each image shows areas that were more active for one type of independent modifier than for the average of the other two types of independent modifier. As can be seen on the left, mental state modifiers, when processed independently, activated classic medial prefrontal areas associated with processing actual mental states in online social tasks. Thus, participants ran simulations of mental states to represent the meanings of the mental state modifiers. The images in the center indicate that classic areas in the left temporal lobe that process motion during actual perception also become active when participants here processed motion modifiers independently. Analogously, the images on the right indicate that classic parahippocampal areas that process locations during actual perception also become active when participants here processed location modifiers independently. Thus, participants ran simulations to represent the meanings of all three modifier types, when they processed the modifiers independently.

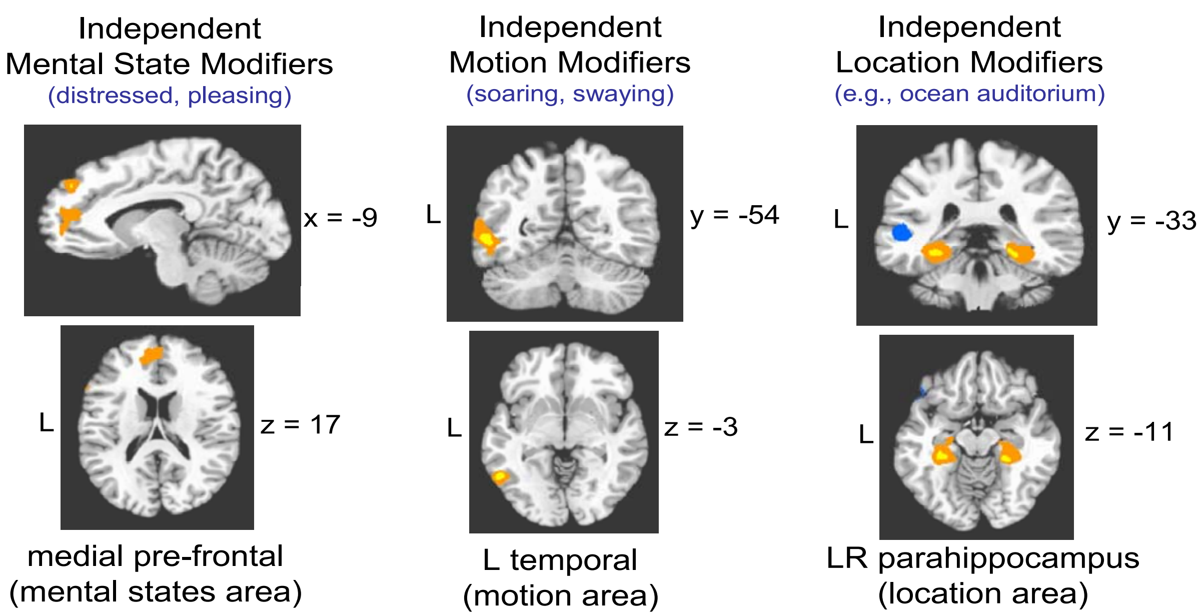


Figure 21. In the fMRI experiment on conceptual combination, brain areas more active for one modifier type than for the other two modifier types on independent trials.

Interestingly, the modality-specific areas in Figure 21 were *not* active when participants processed exactly the same modifiers on combined trials (i.e., participants did not judge the familiarity of the modifier and head noun separately but judged the familiarity of the noun phrase together as a single linguistic unit). As Figure 22 illustrates, the mental state and motion activations observed for processing mental state and motion modifiers independently disappeared. The location activations observed for location modifiers were much reduced, and occurred only on the left, instead of bilaterally, as for the independent trials. This pattern suggests that participants held off committing to particular simulations of the modifiers until processing the head nouns. This makes much sense from a computational standpoint. Because words are so ambiguous in their meaning, and because the meaning of a modifier can be constrained heavily by the meaning of the subsequent head noun, it makes to hold off representing the modifier until the head noun has been comprehended. On receiving "distressed" as a mental state modifier, for example, it's not clear what the meaning of "distressed" will be, given that its meaning can vary widely as a function of the type of person distressed (e.g., mother, child, lawyer). Interestingly, locations are less likely to be affected by the head noun, given that they can be very constraining themselves, often constraining the subsequent head noun more than the head noun constrains them. Thus, it makes sense that some parahippocampal activation remained for the combined location modifiers.

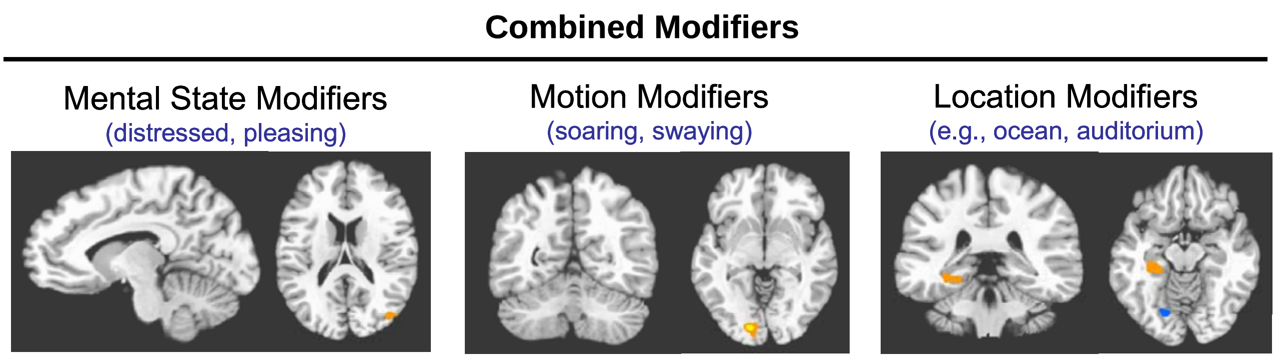


Figure 22. In the fMRI experiment on conceptual combination, brain areas more active for one modifier type than for the other two modifier types on combined trials. The X, Y, and Z coordinates of the slices are identical to those in Figure 21.

Given the general lack of simulation effects for combined modifiers in Figure 22, an interesting question is: What brain areas *do* become active when people process combined modifiers? To answer this question, we subtracted activations for independent modifiers from activations for combined modifiers (across all modifier types together). Figure 23 shows the results. As can be seen, a right hemisphere network became active, including areas in frontal, temporal, and parietal lobes. It is interesting that this is a *right* hemisphere network, given that the materials and task are purely linguistic, which should typically activate the left hemisphere. It is an open question what the role of this right hemisphere network is. Two likely possibilities are as follows. First, other research on comprehension and figurative language has reported right hemisphere activations as people draw inferences from language. This suggests that our participants might be trying to infer what the head noun might be that will follow the modifier. Another possibility, consistent with research on conceptual combination cited earlier, is that participants are trying to set up a relational or situational structure that will eventually incorporate the meanings of both the modifier and head noun after the head noun is presented. Further research is required to resolve this issue.

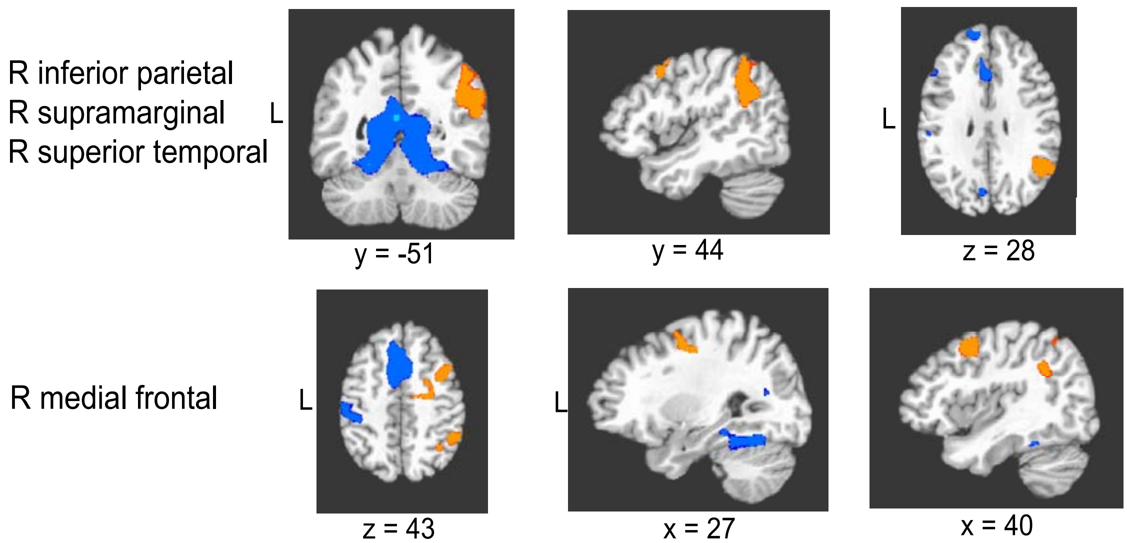


Figure 23. In the fMRI experiment on conceptual combination, brain areas more active for combined modifiers than for independent modifiers, averaged across modifier type.

The last two findings concern the brain areas active while participants processed the head nouns. First, it is important to note that there were no significant activations between the head nouns when processed independently as a function of the modifier that preceded them. This indicates that the semantics of the head nouns were well controlled, such that the set of head nouns following each modifier type was essentially the same. Given this equivalence, it is of interest to ask whether the preceding modifiers affected processing of the head nouns when participants had to combine the modifiers and head nouns on combined trials. Figure 24 shows the results of this analysis. As can be seen, modality-specific simulations occurred for head nouns that followed mental state and location modifiers. Because the content of the head nouns was equivalent (as just discussed), these activations must reflect working memory for the modifiers that remained active while participants processed the head nouns. These activations indicate that simulations of the modifiers were incorporated into the combined representations for the noun phrases constructed as participants processed the head nouns. The apparent lack of motion activation for head nouns following the motion modifiers is probably misleading. As we will see in Figure 25, motion areas were active for head nouns following *all three* types of modifiers. Thus, motion activations for head nouns following the motion modifiers were masked by motion activation following all three modifier types. As will become clear in a moment, participants activated motion simulations for all head nouns in order to simulate entire situations that incorporated the meanings of modifiers and head nouns, with these situational simulations being independent of modifier type.

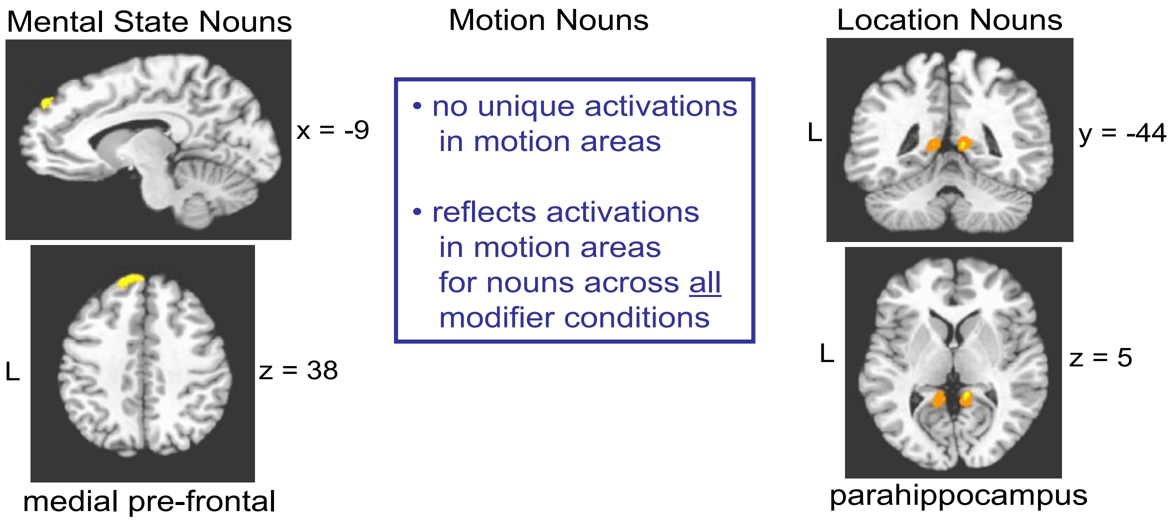


Figure 24. In the fMRI experiment on conceptual combination, brain areas more active for head nouns following one modifier type than for head nouns following the other two modifier types on combined trials.

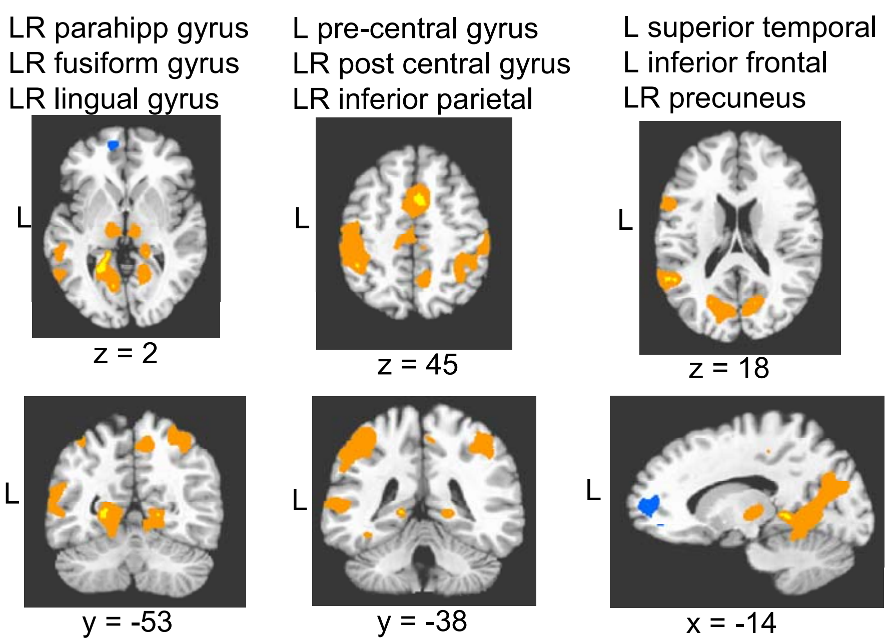


Figure 25. In the fMRI experiment on conceptual combination, brain areas more active for combined head nouns than for independent head nouns, averaged across modifier type.

Figure 25 shows areas that were more active for the head nouns when they were combined than when they were independent (averaged across the type of preceding modifier). Although exactly the same words were processed in the combined and independent conditions, massive differences occurred in activation. The brain was much more active while processing the head nouns when they were combined than when they were independent. The specific areas more active are illuminating. In general, the brain appeared to represent multimodal situations for the combined head nouns but not for the independent head nouns, with many different modalities contributing to these situational representations. As Figure 25 shows, areas that process the physical structure of objects were active (fusiform and lingual gyrus), as were areas that process object motion (temporal gyrus), settings (parahippocampus), action (pre- and post-central gyrus, space (parietal), and imagery (pre-cuneus).

In summary, this study provides intriguing insights into the process of conceptual combination never before observed. When people process modifiers independently, they simulate their meanings. When people process the same modifiers in combinations, however, they hold off committing to their meanings until a head noun has been presented. While waiting for the head noun, people either generate inferences about words likely to follow, or simulate a skeletal situation that could contain both meanings of the modifier and the head noun. Finally, when the head noun arrives on combined trials, it is combined with the meaning of the modifier in a multimodal simulation that represents diverse aspects of a situation, including objects, events, settings, and actions.

3. Further research. The methodology developed in the previous experiment can be used to assess a wide variety of issues in conceptual combination. One of our top priorities in the future is to continue this line of research. Much remains to be learned about the how simulations for individual words are combined and about additional situational information that is inferred. Besides understanding how the meanings of noun phrases are constructed, it will be essential to understand how the more complex conceptual combinations that underlie sentence and text processing are computed. We suspect that the combination of simulations plays a fundamental role in these processes.

IV. Evidence for Situations and Simulation in Abstract Concepts

Two lines of research were developed under this DARPA contract to assess the representation of abstract concepts. One line of research uses a computational linguistics paradigm to assess the role of situations in representing abstract concepts. PSS predicts that abstract concepts capture information about meta-cognitive states and their relations to events during situated action. Thus, situations should play central roles in representing abstract concepts. The second line of research uses an fMRI paradigm to assess an additional prediction from PSS that the meanings of abstract concepts are grounded in simulations of the situations in which these concepts are processed. Both paradigms offer much potential for studying the representation of abstract concepts and for assessing theoretical accounts of them. Each paradigm, along with the results obtained from it to date, is addressed in turn.

A. Scaling Evidence for the Situational Organization of Abstract Concepts

According to PSS, the representations of abstract concepts are grounded in simulations of situations that are distributed across multiple modalities (e.g., Barsalou, 1999; Barsalou & Wiemer-Hastings, 2005). According to this account, these simulations draw heavily on interoceptive states and on the relations of interoceptive states to goal-directed events in the environment. Many abstract concepts appear to provide conceptualizations of meta-cognition as agents pursue situated action. Thus, situations should play central roles in the representations of these concepts.

A prediction that follows is that abstract concepts should be organized according to situations. Abstract concepts used to process the same situation should be associated together, forming a thematic cluster of concepts that play different roles in conceptualizing different but related aspects of the situation. For example, the concepts of *property, ownership,* and *control* form a situational cluster of abstract concepts, because they are relevant to conceptualizing situations that concern personal property, corporate property, government property, etc.

Prior to the work on this project, relatively little research had attempted to identify the organization of abstract concepts. Although extensive research has explored the organization of concrete concepts, the only investigations of abstract concepts have examined small and restricted samples. Thus, our goal here was to examine a large sample of abstract concepts and the organizational structure within it.

Problematically, it is difficult to assess the organizational structure for large samples of concepts using human subjects (e.g., 500 concepts). Although human subjects can be used to assess the organization of small samples, the time and complexity required for assessing the organizational structure of large samples is prohibitive. For this reason, we began searching for other alternatives.

Our search led to tools developed by computational linguists for text analysis. As recent work has shown, these tools can be used to assess the similarity of concepts (e.g., Landauer & Dumais, 1997; Steyvers & Tenenbaum, 2005). Consider Latent Semantic Analysis (LSA). LSA works by first sampling a set of words whose similarities are of interest, where each word is assumed to be associated with a concept (e.g., the word “dog” is associated the concept *dog*). LSA assesses the similarity between the concepts in a sample by creating a vector for each word that represents the frequency of other words that cooccur with it across thousands of texts in an online corpus. After finding context words that surround *dog*, for example, the frequency of these words is stored in a vector, which is later reduced through principle components analysis to make its dimensionality tractable. Much debate exists about what these vectors mean cognitively, but there is no doubt that they are correlated with the similarity of concepts. When the vector for *dog* is compared with the vector for *cat*, these vectors are more similar than those for *dog* and *car*. Interestingly, this similarity reflects their linguistic contexts.

We have used this same general approach to assess the organization of abstract concepts. We did not use LSA, however, because the process of factor analysis introduces structure into the context vectors that is difficult to interpret, and which could potentially distort similarities. Instead, we adopted an approach preferred by some computational linguists that retains the actual sentence contexts of words, rather than reducing them to an arbitrary number of principle components.

It is important to note that our computational analysis does not necessarily reflect how humans organize abstract concepts in their cognitive systems. Instead, our analysis only indicates how abstract concepts cluster according to their linguistic contexts. As described later, however, this analysis offers an intriguing hypothesis about how abstract concepts may be organized in humans. Later research with humans will address this hypothesis.

1. Sampling abstract concepts. To perform these analyses, we first needed to identify abstract concepts. The best source that we could find is the MRC Psycholinguistic Database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm), which includes concreteness ratings for 4,295 words from many different syntactic categories (e.g., nouns, verbs, adjectives). Roughly speaking, these words form a bimodal distribution, with distinct distributions for abstract and. concrete concepts lying on either end of the concreteness continuum, with a large non-modal group of “intermediate” concepts lying between.

Prior to settling on a particular sample of abstracts for further study, we explored a variety of different samples. We eventually settled on a sample of 484 abstract nouns, whose median concreteness rating is 3.3 on a 7 point scale, where 1 is maximally abstract and 7 is maximally concrete. Based on a preliminary scaling analysis, these words differed from words that were intermediate in concreteness. In general, all the words in this sample appear to be bone fide abstract concepts that are not partially abstract and partially concrete. Examples include *aspect, preference*, *justification*, and *responsibility*. In addition, we only included abstract *nouns* that occurs a minimum of 1,000 times in the British National Corpus (BNC). By only including words that had a relatively high frequency, we insured that the words would not be esoteric, and that they would have relatively stable context vectors.

For comparison purposes, we also sampled a cluster of 548 concrete words. Again, these words were only nouns that occurred at least 1,000 times in the BNC. Again, a preliminary scaling analysis determined that they differed from words on the intermediate part of the continuum, having a modal concreteness value of 5.79. Examples include *milk, ear, boat*, and *bed*.

2. The scaling procedure. Each sampled word was then projected onto the BNC such that all sentences containing the word were retrieved. The average number of sentences retrieved per word was 7,636 (the median was 3,928). Context words were then extracted from each of these sentences for a given target word. Similar to how we explored various sampling procedures before settling on a particular sample of abstract concepts, we explored various ways of extracting context words before settling on the context words to extract. We settled on only including open class words from the sentence contexts of the target word (i.e., nouns, verbs, adjectives, adverbs). All other sentence words were discarded and not considered as context. The rationale for this choice was that open class words would be most likely to carry semantic information about the target words and thus be informative about how they should be clustered. Preliminary analyses confirmed this decision.

The words that defined the context vector for each target word resulted from the union of all context words across all target words. Because this union contained 268,040 words, the context vector for each abstract concept contained 268, 040 values. Specifically, the context vector for each target word was the frequency with which each of these context words occurred, where most of the values were 0 (i.e., because most context words did not occur for the target word but occurred for other target words).

Once the context vectors had been formed for the 484 target words, their vectors were submitted to hierarchical clustering. Again, we explored various possibilities before settling on a particular procedure. For the similarity metric, we used the cosine function. For the amalgamation rule, we used Ward's method.

The scaling procedure returned a solution that includes many sensible clusters of abstract concepts at many levels. Figure 26 shows two fragments of the solution (showing the entire solution on a single viewable page is not possible). The similarity between any two words in the solution covaries with the path length between them.

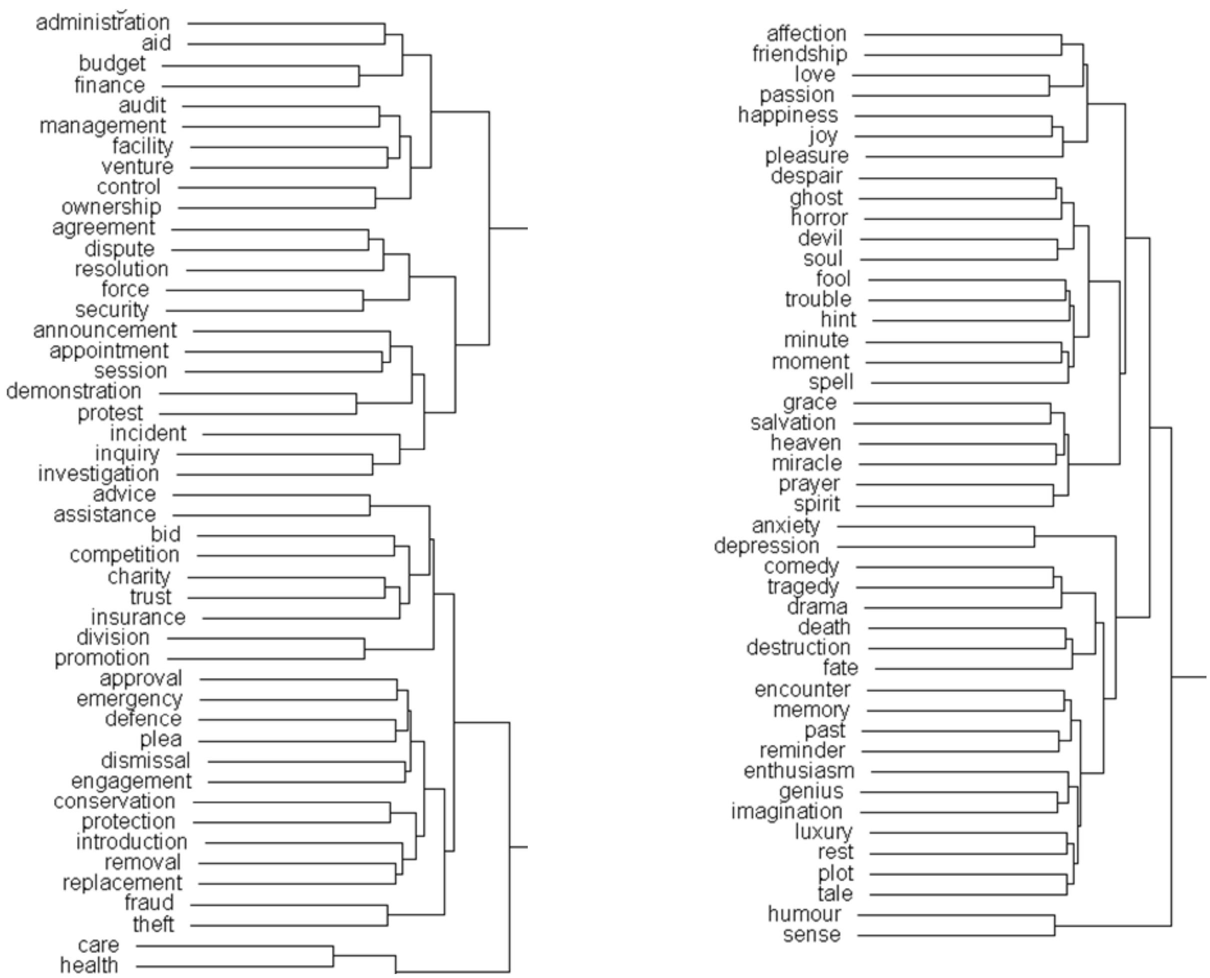


Figure 26. Large fragments of the hierarchical scaling solution for abstract concepts that correspond to two large clusters within it, one related to institutions and the other to personal and interpersonal experience.

Although not all clusters are sensible, the large majority are. As can be seen, the large fragment on the left contained clusters related to institutions, whereas the large fragment on the right contained clusters related to personal and interpersonal experience. Within each large cluster, many small clusters have intuitive interpretations as well. At the top of the right fragment, for example, *affection, friendship, love, passion, happiness, joy,* and *pleasure* form a coherent cluster related to intense positive emotion with other people.

3. Coding analysis. To establish the content of the scaling solution more rigorously, we developed a coding scheme that human judges applied to the clusters at the first four levels of the solution. Table 7 lists the main coding categories in this scheme, which were applied in order from top to bottom. The *first* coding category that fit a cluster in this order was applied with no consideration of later coding categories. This procedure worked against our hypothesis, given that it created a bias against coding a cluster as thematic, which is the type of cluster that we predicted should occur most often, based on the hypothesis that situations organize abstract concepts (i.e., thematic relations organize the components of a situation).

Cluster Type	Cluster Description
Lexical	Cluster elements form a common lexical phrase.
Synonym	Cluster elements are synonyms.
Antonym	Cluster elements are antonyms.
Taxonomic	Cluster elements belong to the same superordinate category, linked by an ISA relation to the superordinate.
Partonymic	Cluster elements belong to the same larger whole, linked by a PART OF relation.
Thematic	Cluster elements co-occur in the same domain, situation, or event, or are connected by any of a wide variety of conceptual relations.
Shared	Cluster elements share collocates, but collocates do not convey any of the specific kinds of commonality in the remaining cluster types. Applies only if no other cluster type can be assigned.

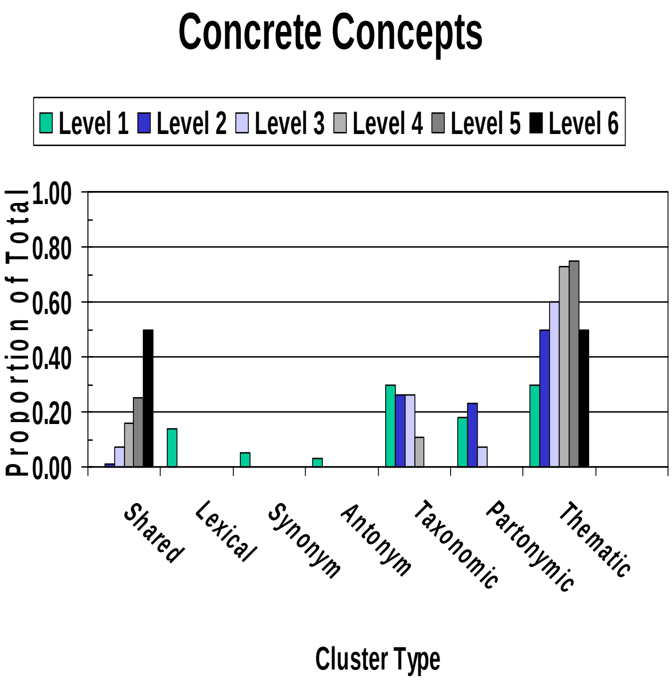
Table 7. The coding scheme used to code types of clusters in the hierarchical scaling solution for abstract concepts.

To assess our hypothesis, we coded the clusters in the scaling solution using the coding scheme shown in Table 7. The first three coding categories are relatively linguistic, capturing clusters that formed lexical compounds, synonyms, or antonyms. The fourth coding category—taxonomic—is the type of organization that researchers assume dominates the organization of concrete concepts. Of interest in the analysis here is whether taxonomic organization plays a significant role in systems of abstract concepts as well. The fifth coding category—partonymic—is another type of organization central for concrete categories (i.e., parts organized into wholes, such as parts of a car).

The sixth coding category—thematic—is typically viewed as the antithesis of taxonomic organization (e.g., Lin & Murphy, 2001). Thematic clusters organize concepts that play different but correlated roles in the same situation, with conceptual relations typically linking them. For example, *hammer, nail,* and *board* are related thematically because each plays a different but interrelated role in the common situation of hammering nails into boards. Our prediction for abstract concepts is that thematic clusters should be highly prevalent in their organization. If people organize abstract concepts together because they are typically processed together in the same situation, then these concepts should frequently fall into thematic clusters.

The final coding category—shared—only applied if no other cluster type could be applied to the cluster. The words in these clusters tended to share collocates, but the collocates did not convey any specific type of commonality.

Figure 27 shows the results of this coding analysis. As predicted, thematic clusters dominate the scaling solutions for abstract concepts. Even for the terminal clusters at the lowest level of the solution (i.e., Level 1 in Figure 14), thematic clusters dominate, occurring 51% of the time. The remaining clusters are taxonomic (20%), lexical (8%), synonyms (8%), and antonyms (5%), indicating that other organization occur for abstract concepts, but at relatively low rates.



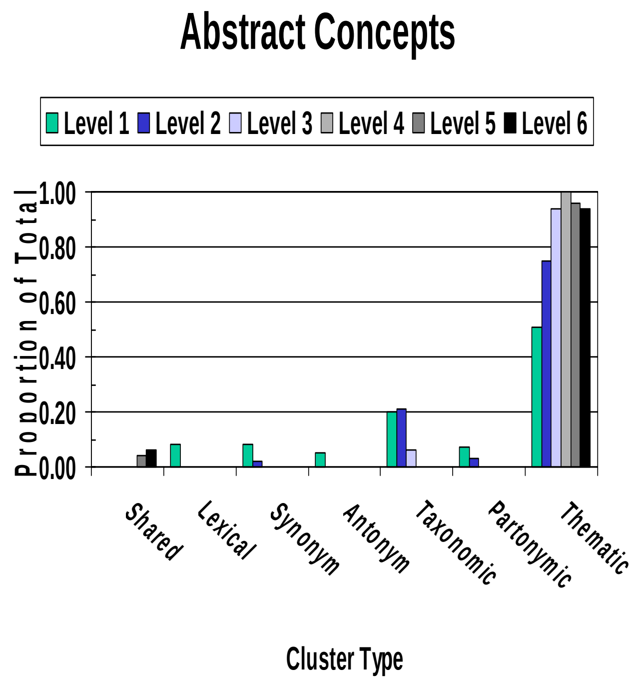


Figure 27. Proportion of each cluster type as function of hierarchical level in the scaling solution, where Level 1 is the terminal level of the solution, and Levels 2, 3, and 4 are increasingly high levels.

At increasingly high levels in the solution, thematic clusters dominate increasingly, occurring at 75% of the time for Level 2, 94% of the time for Level 3, and 100% of the time for Level 4. These results indicate that thematic organization is the dominant organization of abstract concepts, at least as based on text analysis. It is an interesting and open question whether this is the dominant organization in the human cognitive system. Nevertheless, these results tentatively suggest that situations provide the dominant organization of abstract concepts. Clusters of abstract concepts are likely to form based on their situational cooccurrence.

Figure 27 also show results for the concrete concepts. Consistent with thematic clustering being the dominant organization of concepts in the cognitive system, thematic clusters dominate the organization concrete concepts, not just abstract. Nevertheless, taxonomic organization and partonymic organization also play important roles in the organization of concrete concepts. Furthermore, these organizations are more important for concrete concepts than for abstact. Interestingly, taxonomic organization and partonymic organization are most important for concrete concepts at low levels of organization, increasingly giving way to thematic organization at higher levels.

4. Collocates analysis. To further explore clusters in the scaling solution, we wrote a program that returns the contextual collocates of the words in a particular cluster that occurred most often across cluster words. By examining these collocates, we can learn more about why particular abstract words clustered together in this text-based analysis.

Table 8 presents examples of four clusters and some of the collocates that contributed their formation. The words for each cluster are shown in a column on the left. The collocates for the cluster are shown on the right. The collocates are words that occurred frequently in the linguistic contexts individual words in the cluster. Thus, these are the contextual elements that caused the cluster to form. Because the cluster words tended to share these particular collocates, they clustered together. (Note that the collocates were generally shared by *all* words in the cluster, and did not only occur for the word to their immediate left.)

By examining the collocates for a cluster, it is possible to obtain a sense of the situations in which the cluster's words cooccur. For example, words in the top cluster share collocates having to do with financial situations, whereas words in the bottom cluster share collocates having to do interpersonal and family situations. These are likely to be situations in which these clusters of abstract concepts cooccur, such that they become organized together.

Cluster	Frequent and Shared Contextual Collocates
amount **sum** **loan** **cost** **estimate**	percent, government, rate, total, value, number, money, paid, interest, tax
accord **peace** **surrender** **pact** **treaty** **protocol**	government, countries, terms, signed, military, support, forces, agreement, political, states, security, meeting, co-operation
conscience **truth** **courage** **wisdom** **dignity** **pride**	people, life, work, sense, right, government, power, help, human, political, social
anger **grief** **hatred** **guilt** **shame** **mercy** **pity**	people, know, see, love, life, think, felt, feel, great, sense, face, work, death, women, mother, father, eyes

Table 8. Examples of four clusters from the scaling solution for abstract concepts (left) and the contextual collocates that occurred frequently in the linguistic contexts of cluster words (right). Note that the collocates generally occurred for most cluster words and not only for the word to their immediate left.

5. Further research. We have also scaled a carefully selected set of 548 concrete words from the same distribution as the 484 abstract words, using the same sampling principles. Although we have not yet coded the clusters in this solution, there appear to be many, many more taxonomic clusters than in the solution for abstract concepts. Interestingly, however, there appear to be many thematic clusters as well, and the proportion of thematic clusters appears to grow across higher taxonomic levels. Although concrete concepts appear to be organized more taxonomically than abstract concepts, they nevertheless appear to be organized situationally as well.

An important line of research to pursue once the scaling analyses have been completed is to see whether people organize abstract concepts thematically. We have begun planning a series of laboratory experiments to assess this issue.

B. fMRI Evidence for Simulation in the Representation of Abstract Concepts

This sixth and final line of research used fMRI to further address predictions from PSS about the representation of abstract concepts. According to PSS, the representation of an abstract concept is grounded in simulations of the situations in it occurs, with a focus on interoceptive states in the situation and their relation to goal-directed events. Thus, PSS predicts that when people receive the word for an abstract concept (e.g., *arithmetic*), they should simulate the situations in which the concept occurs, especially the relevant interoceptive states and events in these situations. The line of research developed here offers a new way for studying abstract concepts, in general, and for assessing the potential role of simulation, specifically.

This perspective on abstract concepts contrasts sharply with standard views in both the behavioral and neuroscience literatures. In general, other researchers have come to believe that abstract concepts are grounded in language (for reviews, see Binder et al., 2005; Paivio, 1986). This conclusion is based on the well-established findings that, first, mental imagery does not appear to accompany abstract concepts, and second, abstract concepts generally appear to activate classic language areas in the brain, such as left inferior frontal gyrus. Problematically, however, the tasks typically used to measure abstract concepts are often highly linguistic in nature, such as lexical decision and synonym judgment. Thus, strong linguistic effects for abstract concepts could reflect the tasks used. Consistent with this conclusion, when less linguistic tasks are used, or more situational context is provided, abstract concepts behave more like concrete concepts (e.g., Barsalou & Wiemer-Hastings, 2005; Schwaneflugel, 1991). Finally, it is not clear what it means to say that "abstract concepts are grounded in language." If someone tells me words that describe an abstract concept in a language that I do not know, I certainly do not understand the concept. At some point, language about an abstract concept must be grounded in experience, which is the claim of PSS. From this perspective, the meaning of an abstract concept is grounded in the situations in which it occurs. When people need to represent the meanings of abstract concepts, they simulate the respective situations.

An additional methodological factor shaped the paradigm developed here. Typically, neuroimaging studies of concepts use many different concepts and present each concept once each. In studies of abstract concepts, for example, researchers use many different abstract concepts, often presenting each abstract concept just once. Problematically, to detect fMRI activation in a brain area, sufficient signal must aggregate across trials. If different abstract concepts activate somewhat different brain areas, because of variation in their semantics, aggregations may not accumulate that are informative about these semantics. An obvious solution to this problem is to present a small number of abstract concepts many times, so that signals for their semantics can aggregate.

The line of research developed here created a new paradigm that resolves the two methodological problems just described. First, this paradigm forces participants to perform deep processing of abstract concepts, such that processing goes considerably beyond superficial linguistic activation. Second, this paradigm presents a small number of concepts many times, so that signal can aggregate for their semantics. This paradigm also offers an additional innovation that makes it possible to test whether the semantics of abstract concepts are grounded in mental simulation, described shortly. In general, this paradigm offers a new tool for exploring a wide variety of issues surrounding abstract concepts (and also concrete concepts).

1. Materials. Four concepts were examined in this experiment. Two of these concepts were abstract: *convince* and *arithmetic*. These two concepts were selected because their neural localizations can be predicted, to some extent, from other work in the literature: Because *convince* concerns people's mental states (i.e., interoceptions), PSS predicts that processing its meaning should activate areas that represent mental states during social interactions, such as medial prefrontal cortex (e.g., Decety & Sommervile, 2003). Analogously, processing the meaning of *arithmetic* should activate areas used in actual number processing, such as the intraparietal sulcus (e.g., Dahaene et al., 2004). If this new paradigm for localizing concepts is valid, we should see activations for *convince* and *arithmetic* in areas like these.

Two concrete concepts were also included so that we could contrast brain activations for concrete and abstract concepts. These two concepts were *red* and *rolling*. Again, we selected these concepts because likely brain localizations for them have been well established in previous research. Based on Martin et al. (1995), we know that color properties like *red* are processed in posterior brain areas, such as the fusiform gyrus. Based on Beauchamp, Lee, Haxby, and Martin (2003), we know that motion properties like *rolling* are processed in motion areas, such as the superior temporal gyrus. If this new paradigm for localizing concepts is valid, we should see activations for *red* and *rolling* in areas like these.

2. Design. The experiment contained two phases: the priming phase, followed by the localizer phase. The localizer phase allowed us to determine the brain areas that underlie the processing of particular situations relevant to the concepts of interest (*arithmetic*, *convince*, *rolling*, *red*). The priming phase allowed us to assess the semantics of these concepts and whether they involve simulations of relevant situations. Each phase is addressed in turn.

The localizer phase came at the *end* of the experiment so as not to bias performance in the priming phase. During the localizer phase, participants performed blocks of trials (in a blocked design) that assessed the online processing of situational content relevant to the four concepts assessed in the priming phase. During each block of a specific localizer task, participants viewed pictures and had to perform judgment on them, as described in a moment. The pictures were held constant, such that participants processed the same set of pictures in each of the four localizer tasks.

The four localizer tasks were as follows. In the *counting localizer*, participants counted the number of entities in each picture. In the *thoughts localizer*, participants inferred the thoughts of the people in the picture (all pictures contained people). In the *motion localizer*, participants imagined motion within the picture. In the *color localizer*, participants imagined the colors of the objects in the pictures (all pictures were in black and white). As described in a moment, these four localizer task were designed to activate brain areas that we predicted underlie simulations of the four concepts assessed during the priming phase. Specifically, we predicted that *arithmetic* would be represented during the priming phase by simulations that used brain areas active during the counting localizer. Analogously, we predicted that *convince* would be represented by areas active during the thoughts localizer, that *rolling* would be represented by areas active during the motion localizer, and that *red* would be represented by areas active during the color localizer.

The priming phase of the experiment used an event related design. Panel A of Figure 28 illustrates the time course of a trial. On each trial (following random inter-trial jitter), participants viewed the word for one of the four critical concepts for 5 sec. We refer to this 5 sec period, when only the word is present, as the *priming period*, because the word should be priming its meaning during this time (e.g., Stroop, 1935). Following the 5 sec priming period, a photo of a scene appeared for 2.5 sec, and participants' task was to indicate whether the preceding word applied meaningfully to the scene. If *red* had been presented during the priming period, participants indicated "applies" when something in the scene could be red (e.g., a red apple). Similarly, on *rolling*, *convince*, and *arithmetic* trials, respectively, participants indicated "applies" if something in the scene was rolling (e.g., a ball), if one person was trying to convince another of something (e.g., a political rally), or if a person was performing some sort of numerical processing (e.g., measuring a child's height). Panel B of Figure 28 illustrates examples of these pictures.

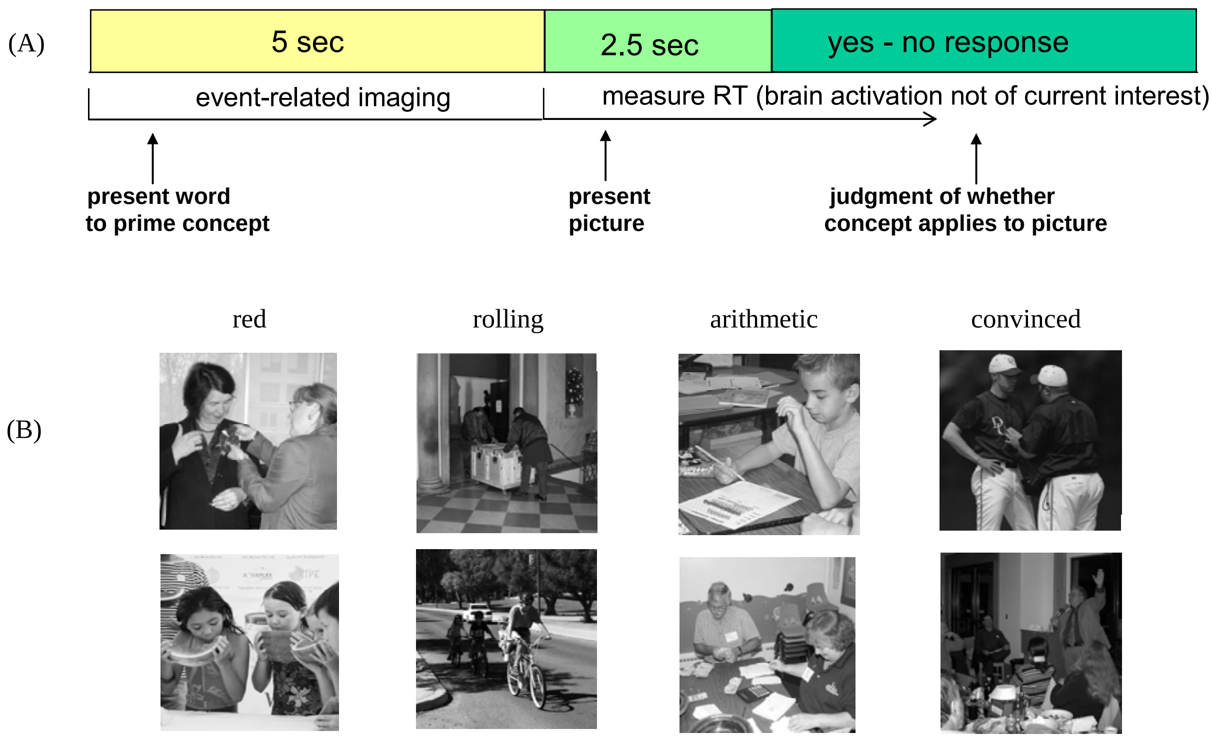


Figure 28. The time course of a trial in the fMRI experiment on abstract concepts (Panel A). Examples of the pictures used for the four concepts (Panel B).

Each of the 4 critical concepts was presented on 36 trials each, applying to the picture on 9 trials and not applying on 27. Each critical concept was presented with the same 36 pictures (not blocked). Thus, different sets of 9 pictures applied to each of the 4 concepts. The same 36 pictures were used for each concept so that different picture sets did not differentially feed back into the priming periods of different concepts (via memory on later trials). The pictures that applied for each concept were selected and scaled to have comparable applicability and visual complexity.

Similar to the previous two fMRI experiments, we used catch trials so that we could deconvolve activations for the priming periods and the pictures. To make these deconvolutions possible, the catch trials presented one of the words for the four concepts *not* followed by a picture. On these trials, the fixation cross reappeared after the priming period, indicating that no picture was coming, and that no response was required. Each of the four words was used equally often on the catch trials.

3. Analysis. Of primary interest were the brain activations that occurred during the priming period, while the critical word was on the screen, before the picture appeared. Brain activations during the picture periods were *not* analyzed. Activations during the priming period were of primary interest because they assess how participants represented the meanings of the four concepts.

Similar to the earlier fMRI experiment on conceptual combination, we used a mask analysis to assess whether the representations of the concepts during the priming period were mental simulations of the relevant situations. Table 9 outlines the steps of this analysis.

Step 1. Create a mask for each localizer.

For each localizer, subtract the activation maps for the other three localizers from its activation map.

The result is the brain areas significantly more active for the localizer than for the other three localizers.

These areas implement the online task that participants perform during the localizer.

The active areas for each localizer will later serve as mask for assessing whether simulations of the localizer represent its respective concept during the priming phase.

Step 2. Deconvolve activations for concepts and pictures during the priming phase.

Using the catch trials, deconvolve the activations for concepts and pictures to produce one activation map for each concept in isolation.

Step 3. Identify activations for each concept within its localizer mask.

Perform the following comparisons for each concept during the priming phase:

Arithmetic – (Convince, Rolling, Red) within the Counting Localizer Mask

Convince – (Arithmetic, Rolling, Red) within the Thoughts Localizer Mask

Rolling – (Red, Arithmetic, Convince) within the Motion Localizer Mask

Red – (Rolling, Arithmetic, Convince) within the Color Localizer Mask

If concepts are represented by mental simulation during the priming phase, then each concept should activate areas in its localizer mask.

Step 4. Identify activations for each concept within <u>the other three localizer</u> masks.

Perform the following three comparisons for each concept during the priming phase:

Arithmetic – (Convince, Rolling, Red) within the Thoughts, Motion, and Color Localizer Masks

Convince – (Arithmetic, Rolling, Red) within the Counting, Motion, and Color Localizer Masks

Rolling – (Arithmetic, Convince, Red) within the Color, Counting, and Thoughts Localizer Masks

Red – (Arithmetic, Convince, Rolling) within the Motion, Counting, and Thoughts Localizer Masks

If concepts are represented by mental simulations during the priming phase, then each concept should <u>not</u> activate areas in the other three localizer masks.

Table 9. Steps of the analysis used to assess whether abstract concepts are represented by simulations of the situations in which they occur.

During Step 1, we created masks of the brain areas active for each localizer. To obtain the localizer mask for counting, for example, we subtracted brain activations for the other three localizer tasks from the activations for the counting localizer task. These masks later allowed us to assess whether participants used simulations of the localizer task when representing concepts during the priming phase of the experiment.

In Step 2, we deconvolved activations for the priming periods from activations for the picture periods during the priming phase of the experiment. The catch trials on which only words were presented (not pictures) made these deconvolutions possible.

Steps 3 and 4 were the critical ones in the analysis. In Step 3, we first identified activations for each concept during the priming phase within the mask for its respective localizer. For example, we identified activations for *arithmetic* within the *counting* mask (Table 9 lists all four critical comparisons). Specifically, activations from the other three concepts (e.g., *convince, rolling, red*) were subtracted from activations for the target concept (e.g., *arithmetic*), but only within the brain regions included in the localizer mask (e.g., counting). This comparison assessed whether representing *arithmetic* during the priming period simulated processes performed during the counting localizer. Similar comparisons were performed for *convince, rolling,* and *red* to see if their activations occurred within their respective localizer masks.

Step 4 assessed whether a target concept (e.g., *arithmetic*) produced activations *outside* its localizer mask (e.g., in the masks for thoughts, motion, and color). If a target concept is represented by simulations of the processes in its localizer task, then it should generally not be represented by processes performed for other localizers.

Analyses of both the mask and priming data were conducted conservatively. For activations to be significant when creating a mask, they had to be significant at $p<.005$ using a spatial correction that took into account the number of voxels tested and the likelihood of contiguous voxels being significant by chance. For activations to be significant during the priming period, within a mask, they had to be significant at $p<.05$, again using a spatial correction. Thus, any voxel significant in the priming periods had to pass two significance tests at a combined level of $p<.00025$ (.005 x .05), plus two spatial corrections (one during the mask analysis, and one during the priming analysis).

4. Primary results. Figure 29 shows the results of primary interest for 14 participants. Each row of the figure represents a localizer task (color, motion, counting, thoughts). The two columns represent the two abstract concepts of primary interest (*arithmetic*, *convince*). The cells within the table represent significant activations for a concept during the priming period (e.g., *arithmetic*) within a localizer mask (e.g., counting). For example, activation in the right cuneus occurred for *arithmetic* within the mask for the counting localizer.

LOCALIZER MASK	PRIMING PERIOD ARITHMETIC				PRIMING PERIOD CONVINCE			
COLOR								
MOTION					L. MTG	-51	-58	10
COUNTING	R. Cuneus	3	-77	37	R. Precuneus	15	-70	52
	R. Precuneus	18	-61	25				
	R. Supramarginal G.	51	-52	30				
	R. Inf. Parietal	33	-39	28				
	R. Inf. Parietal	63	-34	36				
	Corpus Collosum	6	-25	24				
	Thalamus	-17	-16	1				
	L. Postcentral G.	-52	-13	46				
	R. Cingulate	-15	-1	40				
	R. Insula	39	14	7				
	R. Ant. Cingulate	13	32	28				
	R. Frontal	21	38	-4				
THOUGHTS	L. STS/STG	-63	-33	7	R. MTG/STS	57	-61	13
					L. MTG/STS	-54	-58	13
					L. Precuneus	-12	-55	35
					L. Ant. STG/STS	-45	11	-17
					L. SFG, BA 10	-6	62	25

Figure 29. Areas of activation during the priming phase for the two abstract concepts, *arithmetic* and *convince*, within the four localizer masks. Tailarach coordinates (X, Y, Z) for the peak voxel in each active clusters are shown to the right of each significant activation.

Two results within Figure 29 support the PSS account of abstract concepts. First, the activations for both abstract concepts generally do *not* lie in language processing areas. Most importantly, left inferior frontal gyrus is not active for either *arithmetic* or *convince*. Although this has been one of the most frequently active areas for abstract concepts in previous experiments, it was not active here. Nor are other left hemisphere areas active that are often associated with language. This finding strongly suggests that the priming paradigm used here is activating and measuring the semantics of abstract concepts, not just superficial linguistic processing. Previous research appears to have used tasks that require so little processing that they have not activated the semantics of abstract concepts. As the next results illustrate, the priming paradigm here appears to activate their semantics.

First, consider the activations for *arithmetic* in the counting localizer mask. All activations in this cell of Figure 29 can be viewed as areas that were active both during priming periods for *arithmetic* and during the counting localizer. Panel A of Figure 30 illustrates some of these brain areas further. As PSS predicts, the semantics of *arithmetic* shared many activations with the counting localizer. This high overlap strongly suggests that participants used simulations of counting to represent the concept of *arithmetic*, when *arithmetic* was simply cued by a word. Clearly, there is more to doing arithmetic than only counting. Nevertheless, representing the concept of *arithmetic* seemed to draw extensively on the network of brain areas used to perform counting activities, as PSS predicts.

Furthermore, many of these activations lie in posterior brain areas often associated with spatial processing. Other activations lie in frontal areas often associated with motor processing. Most significantly, activations occurred in the intraparietal sulcus, an area frequently associated with mathematical reasoning (i.e., the activation labeled R. Supramarginal G in Figure 29; also see Figure 30A). As this pattern indicates, the priming task activated a much more extensive semantic representation for *arithmetic* than previous theories of abstract concepts would

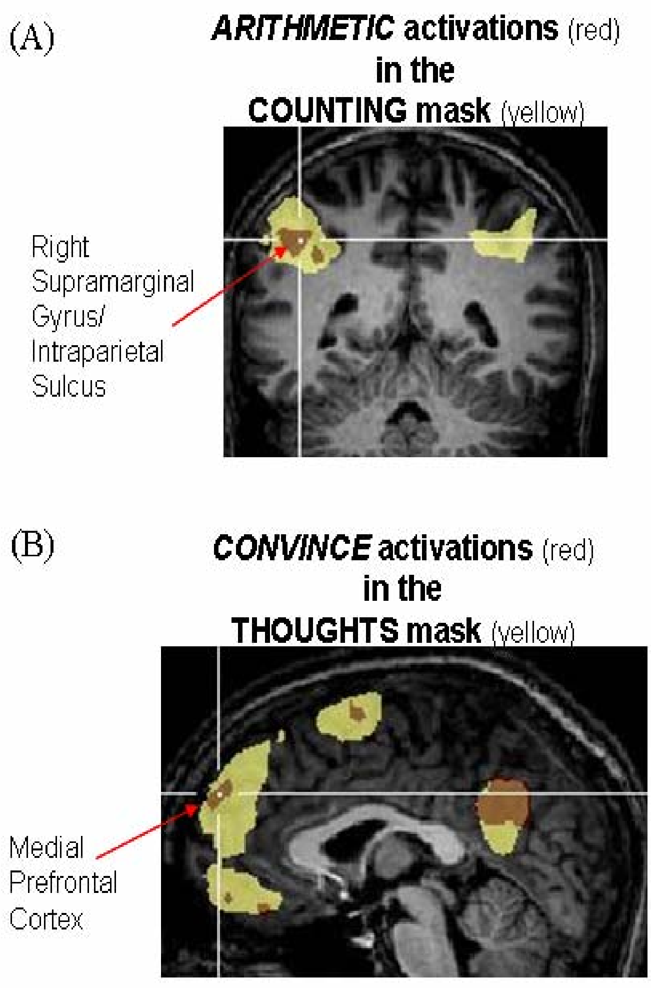


Figure 30. The concept *arithmetic* activates the classic intraparietal sulcus area for mathematical calculation within the counting localizer mask (Panel A). The concept *convince* activates the classic medial prefrontal area for processing mental states within the thoughts localizer mask (Panel B). Mask areas for the localizer are shown in yellow; priming areas for the concept are shown in red.

predict. Because the priming paradigm forced deep processing of *arithmetic,* a rich semantic representation became active, which overlapped extensively with brain areas used for actual counting.

Second, consider the activations for *convince* in the thoughts localizer mask. All activations in this cell of Figure 29 can be viewed as areas that were active both during priming periods for *convince* and during the thoughts localizer. Panel B of Figure 30 illustrates some of these brain areas further. As PSS predicts, the semantics of *convince* shared many activations with the thoughts localizer. This overlap strongly suggests that participants used simulations of thinking to represent the concept of *convince,* when *convince* was simply cued by a word. Clearly, there is more to convincing someone than only thinking. Nevertheless, representing the concept of *convince* seemed to draw extensively on the network of brain areas used to perform thinking activities.

Furthermore, many of these activations lie in posterior brain areas often associated with the processing of visual information during social interaction (e.g., superior temporal gyrus). Other activations lie in frontal areas often associated with mental states (e.g., medial pre-frontal cortex, the area labeled L. SFG BA 10 in Figure 29). Panel B of Figure 30 shows the specific location of this activation. As this pattern indicates, the priming task activated a much more extensive semantic representation for *convince* than previous theories of abstract concepts would predict. Because the priming paradigm forced deep processing of *convince,* a much deeper semantic representation became active, which overlapped extensively with brain areas used for inferring thoughts.

Figure 29 further shows that the activations for an abstract concept were much less likely to lie outside its localizer task than within it. For *arithmetic*, one activation occurred with the thoughts localizer mask. For *convince,* one activation occurred in the motion localizer mask, and one occurred in the count localizer mask. Notably, however, all of these activations are in brain areas associated with motion (STS, MTG) and imagery (precuneus), consistent with the prediction that participants used simulations of events to represent *arithmetic* and *convince*. For *arithmetic* imagining counting motions could activate STS. For *convince*, imagining gestures and facial expressions could activate MTG and the precuneus. Most importantly, however, most of the activations for each concept lay within its localizer mask and not within other localizer masks.

5. Secondary results. Similar results were obtained for the two concrete concepts (*red* and *rolling*), showing that simulations in modality-specific brain areas underlies their semantics during the priming period as well.

We also compared brain activations for the two abstract concepts together versus the two concrete concepts together. Contrary to previous theories, the abstract concepts were not more likely to activate left hemisphere language areas than the concrete concepts. Furthermore, the abstract concepts activated many regions in bilateral posterior areas that process modality-specific information. These results indicate the priming task does indeed go beyond superficial linguistic processing of words to activate a much richer set of semantics than observed in previous research.

Finally, we performed analysis of the activations for the individual concepts over the course of the experiment, and found only minor changes in how they were represented. This finding suggests that the priming paradigm activated a relatively stable set of brain areas across the 36 trials for each concept.

6. Implications and further research. The new paradigm developed here appears to have much potential for measuring the semantic representations of concepts. It allows researchers to measure "deep" representations of a concept that lie beyond superficial activation of word associates (for related results that demonstrate the importance of going beyond superficial linguistic representations, see Barsalou & Solomon, 2004; Glaser, 1992; Kan, Barsalou, Solomon, Minor, & Thompson-Schill, 2003).

In general, this paradigm can be used to identify the brain areas that underlie the semantics of a concept, even when these brain areas cannot be predicted in advance. If researchers want to identify brain areas that represent a particular abstract concept, such as *truth,* they can use this paradigm to do so. By selecting pictures of situations where the concept applies, and then asking participants to verify the word for the concept against the pictures, brain areas that represent "deep" representations of the concept will become active during the priming period.

As the results of this initial experiment illustrate, the semantics of abstract concepts appear to be much more concrete than previously believed, heavily utilizing the brain's modality-specific systems. It will be of interest to see whether future experiments obtain the same result for other abstract concepts. In general, are the semantics of abstract concepts grounded in simulation of the situations in which they occur?

V. Conclusions

A. The Grounding of Symbolic Operations in Simulation

The research performed here offers preliminary support for the proposal that the symbolic operations of predication, conceptual combination, and abstract concepts are grounded in modality-specific simulations. Rather than relying on amodal symbols, predication appears to utilize simulators that represent concept. Conceptual combination similarly appears to rely on the composition of simulations, rather than on the composition of amodal symbols. Abstract concepts appear to be represented by simulations of relevant situations, rather than only by linguistic representations.

It is absolutely essential to state, however, that these conclusions are highly tentative. The research performed here only offers preliminary evidence for the above claims. One or two experiments never demonstrate any major claim definitively. Instead, many years of research are typically required to establish conclusive evidence for claims of this sort (assuming that they are correct). Thus, it will be necessary to first see how the community of researchers responds to these results. Based on their criticisms, observations, and suggestions, it will be necessary to address a variety of issues before stronger conclusions can be reached. It will also be necessary to replicate and extend these findings. And it will be necessary to rule out alternative interpretations of results. It is likely that even more incisive experiments will be developed in the process of publishing results and responding to the research community's reactions. In general, it will take a body of research that is orders of magnitude larger than the research reported here to change how the community thinks about cognitive architecture.

Thus, the results from the research performed under this DARPA contract are highly encouraging but far from conclusive. As a result, caution should be taken in basing any kind of policy on them. On the one hand, these results point in new directions for the design of cognitive architectures. Indeed, it could be tremendously exciting and profitable to implement architectures that these results inspire. On the other hand, we are far from being ready to say that the community *should* build such architectures, because this is how the brain works. Again, many more years of research from many labs will be necessary before we are in a position to make such claims.

Finally, even if the conclusions reached here are correct, their form is likely to evolve considerably as research accumulates and theory evolves. Thus, how these conclusions are conceptualized is likely to change significantly from how they are being conceptualized now.

B. Implications of Symbolic Operations Being Grounded in Simulation

Should the community decide conclusively that symbolic operations are grounded in the brain's modality-specific systems; this conclusion could have considerable impact on artificial intelligence. Rather the performing symbolic operations on amodal symbols in a centralized processor, these operations would be performed in peripheral input-output devices, analogous to how the brain implements symbolic operations in its modality-specific systems. This major change in computational architectures could produce advances that take artificial systems to a new level.

Such a shift could also interact synergistically with the revolution in multi-media processing that has occurred during the past decade. Potential would exist for progressing from using relatively unanalyzed images in digital technology, to having the capability of performing powerful symbolic operations on images. If an artificial system had the ability to learn simulators based on experience (as in Projects 1 and 2 here), these simulators could then be used to interpret regions of images, such that symbolic operations can be performed on them directly, rather than on amodal symbols that stand for them. Essentially, the construct of a simulator offers a natural interface between perception, action, and affective systems on the one hand, and the cognitive system on the other. Indeed, these are not different systems but all one integrated system. Once we have the ability to process images extensively with simulators, the ability of artificial agents to approximate natural agents may increase substantially.

References

Anderson, J.R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Barsalou, L.W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101-140). Cambridge: Cambridge University Press.

Barsalou, L.W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76-121). Cambridge: Cambridge University Press.

Barsalou, L.W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A.C. Collins, S.E. Gathercole, & M.A. Conway (Eds.), *Theories of memory* (pp. 29-101). London: Lawrence Erlbaum Associates.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

Barsalou, L.W. (2003a). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London: Biological Sciences, 358*, 1177-1187.

Barsalou, L.W. (2003b). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, 18, 513-562.

Barsalou, L.W. (2005). Abstraction as dynamic interpretation in perceptual symbol systems. In L. Gershkoff-Stowe & D. Rakison (Eds.), *Building object categories* (389-431). Carnegie Symposium Series. Mahwah, NJ: Erlbaum.

Barsalou, L.W. (2008). Grounded cognition. *Annual Review of Psychology, 59*.

Barsalou, L.W. (in press). Grounding symbolic operations in the brain's modal systems. In G.R. Semin & E.R. Smith (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*. New York: Cambridge University Press.

Barsalou, L.W., Breazeal, C., & Smith, L.B. (2007). Cognition as coordinated non-cognition. *Cognitive Processing*, *8*, 79-91.

Barsalou, L.W., Niedenthal, P.M., Barbey, A., & Ruppert, J. (2003). Social embodiment. In B. Ross (Ed.), *The Psychology of Learning and Motivation,* Vol. 43 (pp. 43-92). San Diego: Academic Press.

Barsalou, L.W., Simmons, W.K., Barbey, A.K., & Wilson, C.D. (2003). *Grounding conceptual knowledge in modality-specific systems. Trends in Cognitive Sciences, 7,* 84-91.

Barsalou, L.W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher and R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129-163). New York: Cambridge University Press.

Beauchamp, M.S., Lee, K.E., Haxby, J.V., & Martin, M. (2003). fMRI Responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience, 15*, 991-1001.

Binder, J.R., Westbury, C.F., McKiernan, K.A., Possing, E.T., & Medler, D.A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience, 17*, 905-917.

Charniak, E., & McDermott, D. (1985). *Introduction to artificial intelligence*. Reading, MA: Addison-Wesley.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General, 132*, 163-201.

Dehaene, S., Molko, N., Cohen, L., & Wilson, A.J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology, 14,* 218-224.

Damasio, A.R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33,* 25-62.

Decety, J., & Sommerville, J.A. (2003). Shared representations between self and other: A social cognitive neuroscience view. *Trends in Cognitive Sciences, 12*, 527-533.

Estes, Z., Verges, M., & Barsalou, L.W. (submitted). Head up, foot down: Object words orient

attention to the object’s typical location.
Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General, 120*, 339-357.
Fodor, J.A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
Gagne, C.I., & Shoben, E.J. (2002). Priming relations in ambiguous noun-noun combinations. *Memory & Cognition*, *30*, 637-646.
Glaser, W.R. (1992). Picture naming. *Cognition,* 42, 61-106.
Kan, I.P., Barsalou, L.W., Solomon, K.O., Minor, J.K., & Thompson-Schill, S.L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology*, *20*, 525-540.
Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's Problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.
Levi, J. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.
Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General, 130*, 3-28.
Marslen-Wilson, W.D., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1-71.
Martin, A. (2001). Functional neuroimaging of semantic memory. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 153-186). Cambridge, MA: MIT Press.
Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58,* 25-45.
Martin, A., Haxby, J.V., Lalonde, F.M., Wiggs, C.L., & Ungerleider, L.G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science, 270*, 102-105.
McRae, K., de Sa, V.R., & Seidenberg, M.S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126,* 99-130.
Meier, B.P., & Robinson, M.D. (2004). Why the sunny side is up? Associations between affect and vertical position. *Psychological Science, 15,* 243-247.
Niedenthal, P.M., Barsalou, L.W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9,* 184-211.
Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
Paivio, A. (1986). *Mental representations: A dual coding approach.* New York: Oxford University Press.
Pollack, J. (1990). Recursive distributed representations. *Artificial Intelligence*, *46*, 77-105.
Prinz, J. (2002). *Furnishing the mind: Concepts and their perceptual basis.* Cambridge, MA: MIT.
Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22,* 253-336.
Pylyshyn, Z.W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80,* 1-24.
Pylyshyn, Z.W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
Ollinger, J.M., Shulman, G.L., & Corbetta, M. 2001a). Separating processes within a trial in event-related functional MRI: I. The method. *Neuroimage, 13*, 210-217.
Ollinger, J.M., Shulman, G.L., & Corbetta, M. (2001b). Separating processes within a trial in event-related functional MRI: II. Analysis. *Neuroimage, 13,* 218-229.
O'Reilly, R.C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.

Rumelhart, D.E., & McClelland, J.L., & the PDP Research Group (Eds.) *Parallel distributed processing. Explorations in the microstructure of cognition, Vol. 1: Foundations*. Cambridge, MA: MIT Press.

Schubert, T. (2005). Your Highness: Vertical positions as perceptual symbols of power. *Journal of Personality and Social Psychology, 89*, 1-21.

Schwanenflugel, P.J. (1991). Why are abstract concepts hard to understand? In P.J. Schwanenflugel (Ed.), *The psychology of word meaning*, (pp. 223-250). Mahwah, NJ: Erlbaum.

Schyns, P.G., Goldstone, R.L., & Thibaut, J.P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21*, 1-54.

Simmons, W.K., & Barsalou, L.W. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology, 20*, 451-486.

Smith, L., & Gasser, M. (2005). The development of embodied cognition: Six lessons from babies. *Artificial Life, 11*, 13-29.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence, 46*, 159-216.

Solomon, K.O., & Barsalou, L.W. (2004). Perceptual simulation in property verification. *Memory & Cognition, 32*, 244-259.

Steyvers, M., & Tenenbaum, J. (2005). The Large Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science, 29*, 41-78.

Stroop, J.R. (1935). Studies on interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Thelen, E. (2000). Grounded in the world: Developmental origins of the embodied mind. *Infancy, 1*, 3-30.

Thompson-Schill, S.L. (2003). Neuroimaging studies of semantic memory: inferring "how" from "where". *Neuropsychologia, 41*, 280-292.

van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science, 14*, 355-384.

Wiemer-Hastings, K., Krug, J., & Xu, X. (2001). Imagery, context availability, contextual constraint, and abstractness. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, 1134-1139. Mahwah, NJ: Erlbaum.

Wu, L., & Barsalou, L.W. (in preparation). Occlusion in conceptual combination.

Zwaan, R.A., & Madden, C.J. (2005). Embodied sentence comprehension. In D. Pecher and R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 224-245). New York: Cambridge University Press.